

List of Probability Distributions

compiled by Mark Herkommer

Contents

0.1	List of probability distributions	1
0.1.1	Discrete distributions	1
0.1.2	Continuous distributions	4
0.1.3	Mixed discrete/continuous distributions	13
0.1.4	Joint distributions	13
0.1.5	Non-numeric distributions	14
0.1.6	Miscellaneous distributions	14
0.1.7	See also	14
1	Discrete Distributions - With Finite Support	16
1.1	Bernoulli distribution	16
1.1.1	Properties	16
1.1.2	Related distributions	17
1.1.3	See also	18
1.1.4	Notes	18
1.1.5	References	18
1.1.6	External links	18
1.2	Rademacher distribution	18
1.2.1	Mathematical formulation	18
1.2.2	van Zuijlen's bound	19
1.2.3	Bounds on sums	19
1.2.4	Applications	19
1.2.5	Related distributions	20
1.2.6	References	20
1.3	Binomial distribution	20
1.3.1	Specification	20
1.3.2	Example	22
1.3.3	Mean and variance	23
1.3.4	Mode	23
1.3.5	Median	23

1.3.6	Covariance between two binomials	24
1.3.7	Related distributions	24
1.3.8	Confidence intervals	27
1.3.9	Generating binomial random variates	28
1.3.10	Tail Bounds	28
1.3.11	See also	29
1.3.12	References	29
1.3.13	External links	30
1.4	Beta-binomial distribution	30
1.4.1	Motivation and derivation	31
1.4.2	Moments and properties	31
1.4.3	Point estimates	32
1.4.4	Further Bayesian considerations	34
1.4.5	Shrinkage factors	36
1.4.6	Related distributions	36
1.4.7	See also	36
1.4.8	References	37
1.4.9	External links	37
1.5	Degenerate distribution	37
1.5.1	Constant random variable	37
1.6	Hypergeometric distribution	38
1.6.1	Definition	38
1.6.2	Combinatorial identities	39
1.6.3	Application and example	39
1.6.4	Symmetries	40
1.6.5	Hypergeometric test	40
1.6.6	Order of draws	41
1.6.7	Related distributions	41
1.6.8	Multivariate hypergeometric distribution	41
1.6.9	See also	42
1.6.10	Notes	42
1.6.11	References	42
1.6.12	External links	42
1.7	Poisson binomial distribution	43
1.7.1	Mean and variance	43
1.7.2	Probability mass function	43
1.7.3	Entropy	44
1.7.4	See also	44

1.7.5	References	44
1.8	Fisher's noncentral hypergeometric distribution	45
1.8.1	Univariate distribution	45
1.8.2	Multivariate distribution	46
1.8.3	Applications	47
1.8.4	Software available	47
1.8.5	See also	47
1.8.6	References	47
1.9	Wallenius' noncentral hypergeometric distribution	48
1.9.1	Univariate distribution	48
1.9.2	Multivariate distribution	50
1.9.3	Complementary Wallenius' noncentral hypergeometric distribution	51
1.9.4	Software available	51
1.9.5	See also	51
1.9.6	References	51
1.10	Benford's law	52
1.10.1	Mathematical statement	52
1.10.2	Example	53
1.10.3	History	53
1.10.4	Explanations	53
1.10.5	Applications	55
1.10.6	Statistical tests	56
1.10.7	Generalization to digits beyond the first	57
1.10.8	Tests with common distributions	57
1.10.9	Distributions known to obey Benford's law	58
1.10.10	Distributions known to not obey Benford's law	58
1.10.11	Criteria for distributions expected and not expected to obey Benford's Law	58
1.10.12	Moments	59
1.10.13	See also	59
1.10.14	References	59
1.10.15	Further reading	61
1.10.16	External links	62
2	Continuous Distributions - Supported on semi-infinite intervals, usually $[0, \infty)$	72
2.1	Beta prime distribution	72
2.1.1	Generalization	73
2.1.2	Properties	74
2.1.3	Related distributions	74
2.1.4	Notes	74

2.1.5	References	74
2.2	Birnbaum–Saunders distribution	74
2.2.1	Theory	75
2.2.2	Properties	75
2.2.3	Probability density function	76
2.2.4	Standard fatigue life distribution	76
2.2.5	Cumulative distribution function	76
2.2.6	Quantile function	77
2.2.7	External links	77
2.2.8	References	77
2.3	Chi distribution	77
2.3.1	Characterization	78
2.3.2	Properties	78
2.3.3	Related distributions	79
2.3.4	See also	80
2.3.5	External links	80
2.4	Chi-squared distribution	80
2.4.1	Definition	80
2.4.2	Introduction to the chi-squared distribution	81
2.4.3	Characteristics	83
2.4.4	Relation to other distributions	86
2.4.5	Generalizations	88
2.4.6	Applications	88
2.4.7	Table of χ^2 value vs p -value	89
2.4.8	History and name	89
2.4.9	See also	89
2.4.10	References	90
2.4.11	Further reading	90
2.4.12	External links	91
2.5	Dagum distribution	91
2.5.1	Definition	91
2.5.2	References	92
2.5.3	External links	92
2.6	Exponential distribution	92
2.6.1	Characterization	92
2.6.2	Properties	93
2.6.3	Parameter estimation	98
2.6.4	Generating exponential variates	99

2.6.5	Related distributions	100
2.6.6	Applications of exponential distribution	102
2.6.7	See also	104
2.6.8	References	104
2.6.9	External links	105
2.7	F-distribution	105
2.7.1	Definition	105
2.7.2	Characterization	107
2.7.3	Generalization	107
2.7.4	Related distributions and properties	108
2.7.5	See also	108
2.7.6	References	109
2.7.7	External links	109
2.8	Fisher's z-distribution	109
2.8.1	Related Distribution	110
2.8.2	References	110
2.8.3	External links	110
2.9	Folded normal distribution	110
2.9.1	Differential equations	111
2.9.2	Related distributions	111
2.9.3	See also	111
2.9.4	External links	111
2.9.5	References	111
2.10	Fréchet distribution	112
2.10.1	Characteristics	112
2.10.2	Applications	113
2.10.3	Related distributions	113
2.10.4	Properties	113
2.10.5	See also	113
2.10.6	References	114
2.10.7	Publications	114
2.10.8	External links	114
2.11	Gamma distribution	114
2.11.1	Characterization using shape k and scale θ	115
2.11.2	Characterization using shape α and rate β	115
2.11.3	Properties	116
2.11.4	Parameter estimation	119
2.11.5	Generating gamma-distributed random variables	120

2.11.6	Related distributions	122
2.11.7	Applications	123
2.11.8	Notes	124
2.11.9	References	125
2.11.10	External links	125
2.12	Generalized gamma distribution	125
2.12.1	Characteristics	125
2.12.2	Moments	126
2.12.3	Kullback-Leibler divergence	126
2.12.4	Software implementation	126
2.12.5	See also	126
2.12.6	References	126
2.13	Generalized Pareto distribution	127
2.13.1	Definition	127
2.13.2	Characterization	127
2.13.3	Characteristic and Moment Generating Functions	128
2.13.4	Special cases	128
2.13.5	Generating generalized Pareto random variables	128
2.13.6	See also	128
2.13.7	Notes	129
2.13.8	References	129
2.13.9	External links	129
2.14	Gamma/Gompertz distribution	129
2.14.1	Specification	129
2.14.2	Properties	130
2.14.3	Related distributions	130
2.14.4	See also	130
2.14.5	Notes	130
2.14.6	References	130
2.15	Gompertz distribution	131
2.15.1	Specification	131
2.15.2	Properties	132
2.15.3	Related distributions	132
2.15.4	See also	132
2.15.5	Notes	133
2.15.6	References	133
2.16	Half-normal distribution	133
2.16.1	Parameter estimation	135

2.16.2	Related distributions	135
2.16.3	External links	135
2.16.4	References	135
2.17	Hotelling's T -squared distribution	135
2.17.1	The distribution	135
2.17.2	Hotelling's T -squared statistic	136
2.17.3	Hotelling's two-sample T -squared statistic	137
2.17.4	See also	138
2.17.5	References	139
2.17.6	External links	139
2.18	Inverse Gaussian distribution	139
2.18.1	Properties	139
2.18.2	Relationship with Brownian motion	140
2.18.3	Maximum likelihood	141
2.18.4	Generating random variates from an inverse-Gaussian distribution	141
2.18.5	Related distributions	142
2.18.6	History	142
2.18.7	Software	142
2.18.8	See also	142
2.18.9	Notes	143
2.18.10	References	143
2.18.11	External links	143
2.19	Lévy distribution	143
2.19.1	Definition	143
2.19.2	Related distributions	145
2.19.3	Random sample generation	145
2.19.4	Applications	145
2.19.5	Footnotes	145
2.19.6	Notes	145
2.19.7	References	146
2.19.8	External links	146
2.20	Log-Cauchy distribution	146
2.20.1	Characterization	146
2.20.2	Properties	147
2.20.3	Estimating parameters	147
2.20.4	Uses	147
2.20.5	References	147
2.21	Log-Laplace distribution	148

2.21.1	Characterization	148
2.21.2	References	149
2.22	Log-logistic distribution	149
2.22.1	Characterisation	149
2.22.2	Properties	150
2.22.3	Applications	151
2.22.4	Related distributions	152
2.22.5	See also	152
2.22.6	References	152
2.23	Log-normal distribution	153
2.23.1	Notation	153
2.23.2	Characterization	154
2.23.3	Properties	155
2.23.4	Occurrence	157
2.23.5	Maximum likelihood estimation of parameters	158
2.23.6	Multivariate log-normal	158
2.23.7	Related distributions	159
2.23.8	Similar distributions	160
2.23.9	See also	160
2.23.10	Notes	160
2.23.11	References	162
2.23.12	Further reading	162
2.23.13	External links	162
2.24	Lomax distribution	162
2.24.1	Characterization	162
2.24.2	Relation to the Pareto distribution	163
2.24.3	Relation to generalized Pareto distribution	163
2.24.4	Relation to q-exponential distribution	163
2.24.5	Non-central moments	163
2.24.6	See also	163
2.24.7	References	163
2.25	Geometric stable distribution	164
2.25.1	Characteristics	164
2.25.2	Relationship to the stable distribution	164
2.25.3	References	165
2.26	Nakagami distribution	165
2.26.1	Characterization	166
2.26.2	Parameter estimation	166

2.26.3	Generation	166
2.26.4	History and applications	167
2.26.5	References	167
2.27	Pareto distribution	167
2.27.1	Definition	167
2.27.2	Properties	168
2.27.3	Generalized Pareto distributions	169
2.27.4	Applications	171
2.27.5	Relation to other distributions	171
2.27.6	Lorenz curve and Gini coefficient	173
2.27.7	Parameter estimation	173
2.27.8	Graphical representation	174
2.27.9	Random sample generation	174
2.27.10	Variants	174
2.27.11	See also	175
2.27.12	Notes	175
2.27.13	References	176
2.27.14	External links	177
2.28	Pearson distribution	177
2.28.1	History	177
2.28.2	Definition	178
2.28.3	Particular types of distribution	178
2.28.4	Relation to other distributions	183
2.28.5	Applications	184
2.28.6	Notes	184
2.28.7	Sources	184
2.29	Phase-type distribution	185
2.29.1	Definition	185
2.29.2	Characterization	186
2.29.3	Special cases	186
2.29.4	Examples	186
2.29.5	Generating samples from phase-type distributed random variables	188
2.29.6	Approximating other distributions	189
2.29.7	Fitting a phase type distribution to data	189
2.29.8	See also	189
2.29.9	References	189
2.30	Rayleigh distribution	190
2.30.1	Definition	190

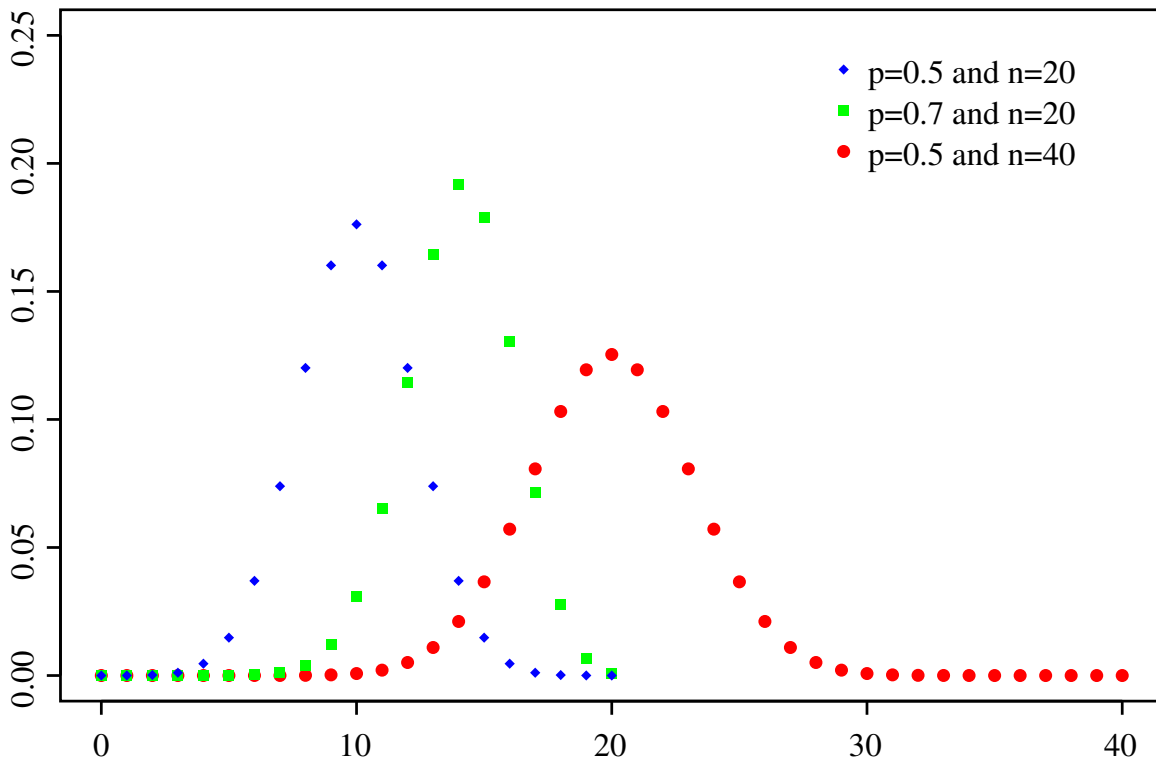
2.30.2	Relation to random vector lengths	191
2.30.3	Properties	191
2.30.4	Parameter estimation	192
2.30.5	Generating random variates	193
2.30.6	Related distributions	193
2.30.7	Applications	194
2.30.8	Proof of correctness – Unequal variances	194
2.30.9	See also	195
2.30.10	References	195
2.31	Rayleigh mixture distribution	195
2.31.1	See also	195
2.31.2	References	196
2.32	Rice distribution	196
2.32.1	Characterization	196
2.32.2	Properties	196
2.32.3	Related distributions	198
2.32.4	Limiting cases	198
2.32.5	Parameter estimation (the Koay inversion technique)	198
2.32.6	Applications	199
2.32.7	See also	199
2.32.8	Notes	199
2.32.9	References	200
2.32.10	External links	200
2.33	Shifted Gompertz distribution	201
2.33.1	Specification	201
2.33.2	Properties	201
2.33.3	Related distributions	202
2.33.4	See also	202
2.33.5	References	202
2.34	Type-2 Gumbel distribution	202
2.34.1	See also	203
2.35	Weibull distribution	203
2.35.1	Definition	203
2.35.2	Properties	204
2.35.3	Weibull plot	206
2.35.4	Related distributions	207
2.35.5	See also	208
2.35.6	References	209

2.35.7 Bibliography	209
2.35.8 External links	210
3 Text and image sources, contributors, and licenses	227
3.1 Text	227
3.2 Images	232
3.3 Content license	236

0.1 List of probability distributions

Many probability distributions are so important in theory or applications that they have been given specific names.

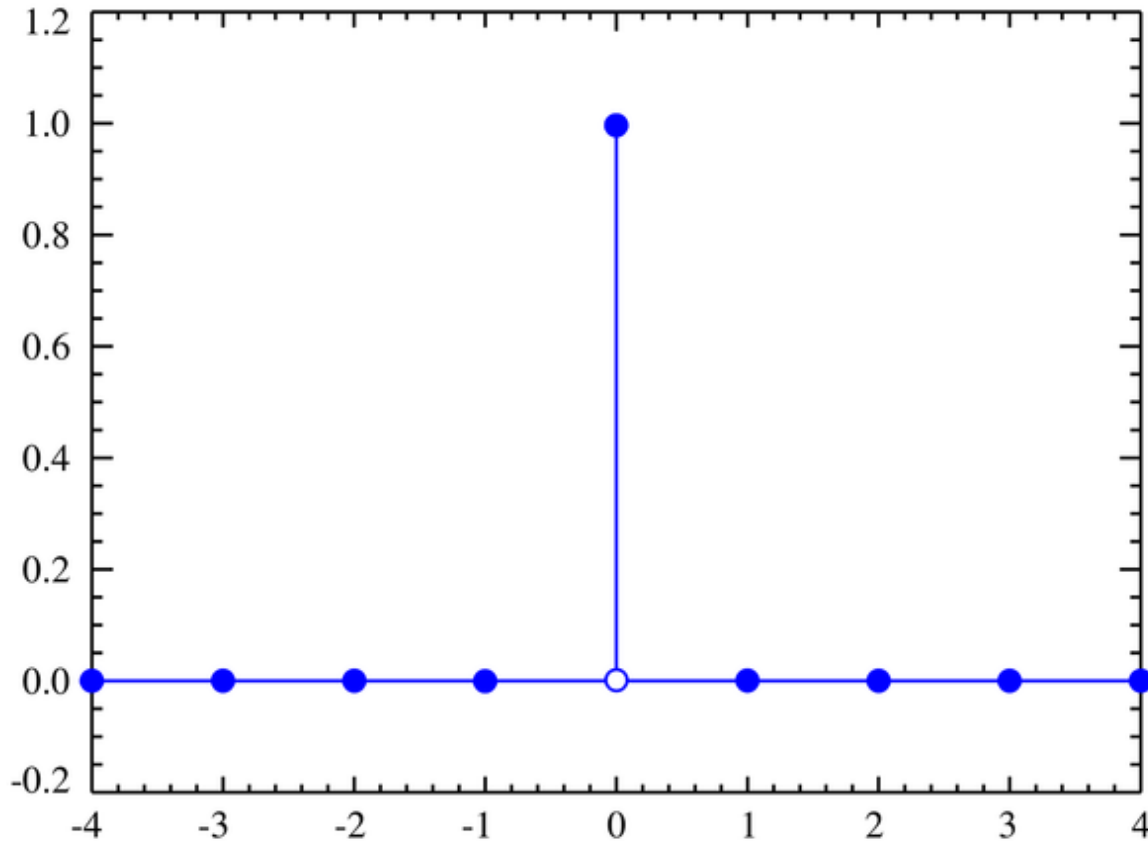
0.1.1 Discrete distributions



Binomial distribution

With finite support

- The **Bernoulli distribution**, which takes value 1 with probability p and value 0 with probability $q = 1 - p$.
- The **Rademacher distribution**, which takes value 1 with probability $1/2$ and value -1 with probability $1/2$.
- The **binomial distribution**, which describes the number of successes in a series of independent Yes/No experiments all with the same probability of success.
- The **beta-binomial distribution**, which describes the number of successes in a series of independent Yes/No experiments with heterogeneity in the success probability.
- The **degenerate distribution** at x_0 , where X is certain to take the value x_0 . This does not look random, but it satisfies the definition of **random variable**. This is useful because it puts deterministic variables and random variables in the same formalism.
- The **discrete uniform distribution**, where all elements of a finite set are equally likely. This is the theoretical distribution model for a balanced coin, an unbiased die, a casino roulette, or the first card of a well-shuffled deck.

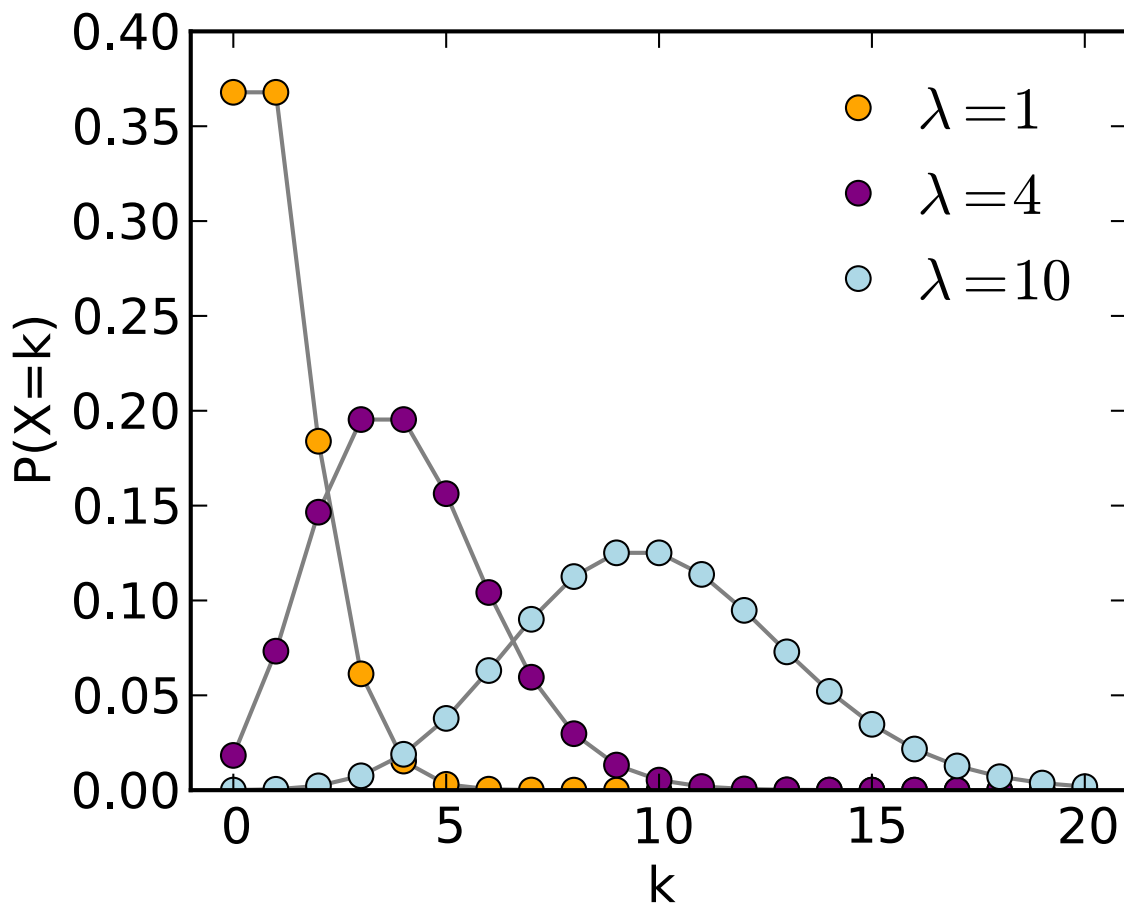


Degenerate distribution

- The **hypergeometric distribution**, which describes the number of successes in the first m of a series of n consecutive Yes/No experiments, if the total number of successes is known. This distribution arises when there is no replacement.
- The **Poisson binomial distribution**, which describes the number of successes in a series of independent Yes/No experiments with different success probabilities.
- Fisher's noncentral hypergeometric distribution
- Wallenius' noncentral hypergeometric distribution
- Benford's law, which describes the frequency of the first digit of many naturally occurring data.

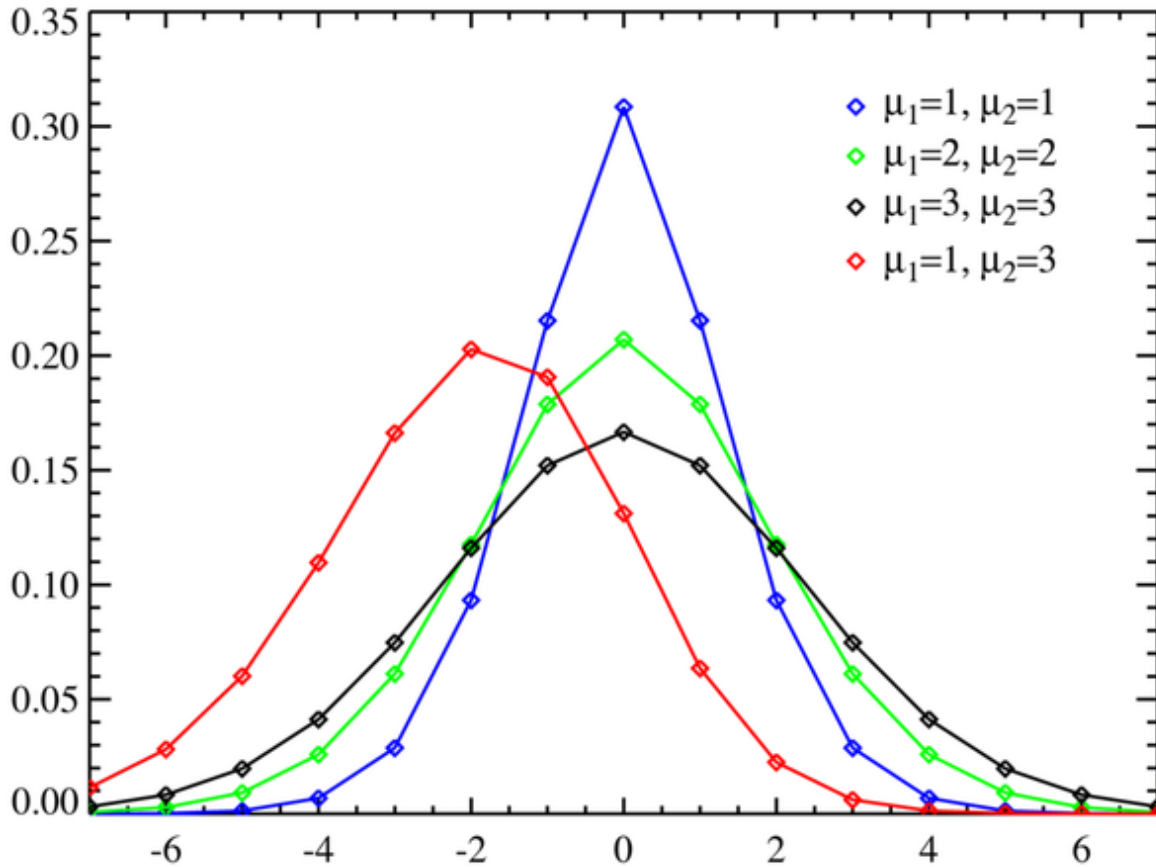
With infinite support

- The **beta negative binomial distribution**
- The **Boltzmann distribution**, a discrete distribution important in statistical physics which describes the probabilities of the various discrete energy levels of a system in **thermal equilibrium**. It has a continuous analogue. Special cases include:
 - The **Gibbs distribution**
 - The **Maxwell–Boltzmann distribution**



Poisson distribution

- The Borel distribution
- The extended negative binomial distribution
- The extended hypergeometric distribution
- The generalized log-series distribution
- The **geometric distribution**, a discrete distribution which describes the number of attempts needed to get the first success in a series of independent Bernoulli trials, or alternatively only the number of losses before the first success (i.e. one less).
- The logarithmic (series) distribution
- The **negative binomial distribution** or Pascal distribution a generalization of the geometric distribution to the n th success.
- The **parabolic fractal distribution**
- The **Poisson distribution**, which describes a very large number of individually unlikely events that happen in a certain time interval. Related to this distributions are a number of other distributions: the **displaced Poisson**, the hyper-Poisson, the general Poisson binomial and the Poisson type distributions.
 - The Conway–Maxwell–Poisson distribution, a two-parameter extension of the Poisson distribution with an adjustable rate of decay.



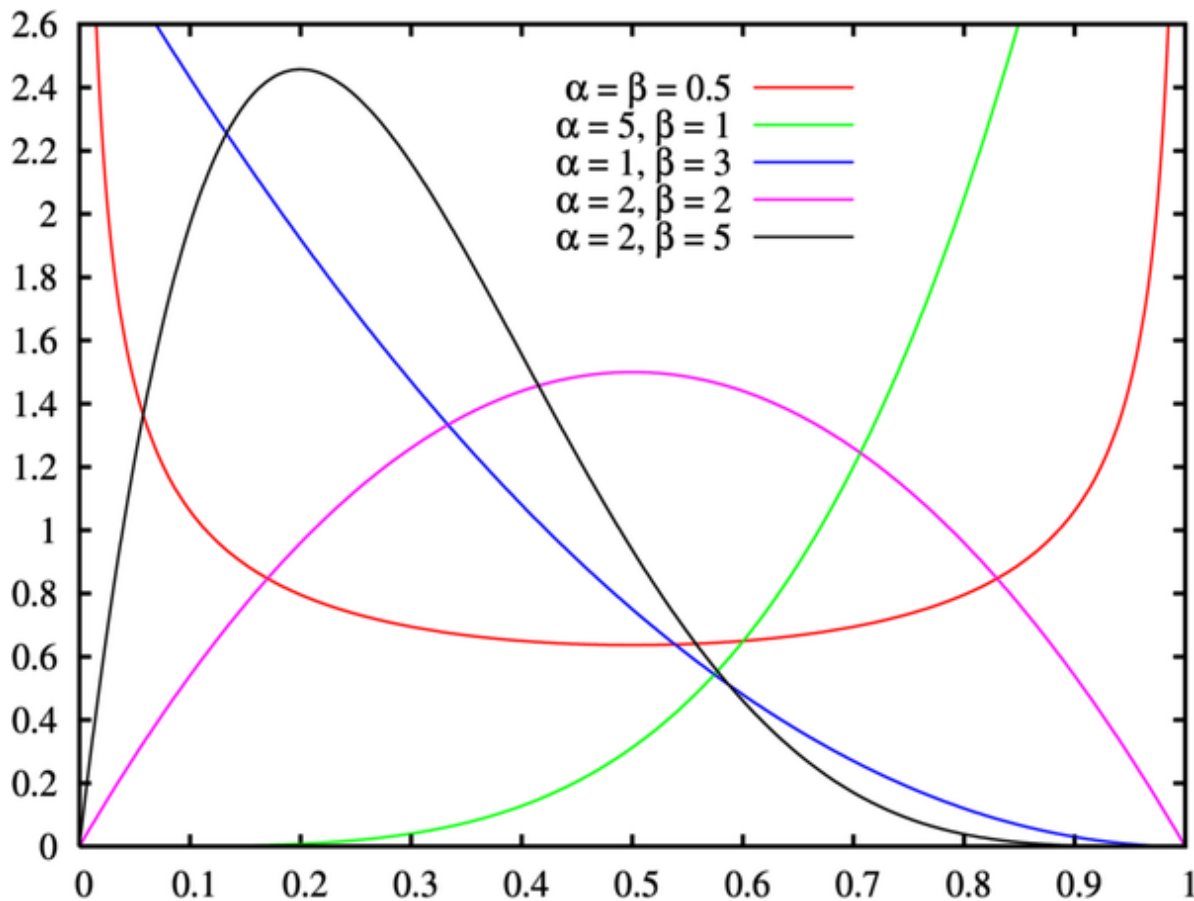
Skellam distribution

- The Zero-truncated Poisson distribution, for processes in which zero counts are not observed
- The Polya–Eggenberger distribution
- The Skellam distribution, the distribution of the difference between two independent Poisson-distributed random variables.
- The skew elliptical distribution
- The Yule–Simon distribution
- The zeta distribution has uses in applied statistics and statistical mechanics, and perhaps may be of interest to number theorists. It is the Zipf distribution for an infinite number of elements.
- Zipf's law or the Zipf distribution. A discrete power-law distribution, the most famous example of which is the description of the frequency of words in the English language.
- The Zipf–Mandelbrot law is a discrete power law distribution which is a generalization of the Zipf distribution.

0.1.2 Continuous distributions

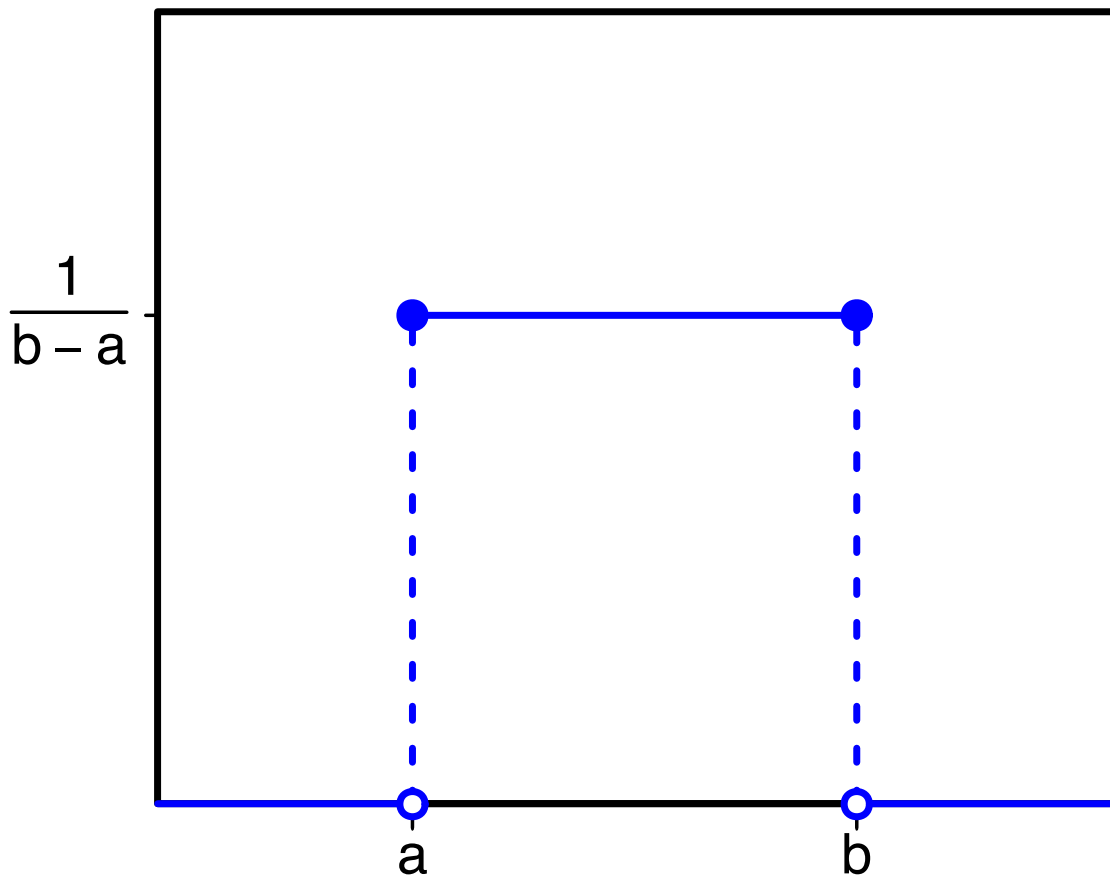
Supported on a bounded interval

- The Arcsine distribution on $[a,b]$, which is a special case of the Beta distribution if $a=0$ and $b=1$.



Beta distribution

- The **Beta distribution** on $[0,1]$, a family of two-parameter distributions with one mode, of which the uniform distribution is a special case, and which is useful in estimating success probabilities.
- The **Logitnormal distribution** on $(0,1)$.
- The **Dirac delta function** although not strictly a function, is a limiting form of many continuous probability functions. It represents a *discrete* probability distribution concentrated at 0 — a **degenerate distribution** — but the notation treats it as if it were a continuous distribution.
- The **continuous uniform distribution** on $[a,b]$, where all points in a finite interval are equally likely.
 - The **rectangular distribution** is a uniform distribution on $[-1/2, 1/2]$.
- The **Irwin–Hall distribution** is the distribution of the sum of n i.i.d. $U(0,1)$ random variables.
- The **Bates distribution** is the distribution of the mean of n i.i.d. $U(0,1)$ random variables.
- The **Kent distribution** on the three-dimensional sphere.
- The **Kumaraswamy distribution** is as versatile as the Beta distribution but has simple closed forms for both the cdf and the pdf.
- The **logarithmic distribution** (continuous)
- The **Marchenko–Pastur distribution** is important in the theory of **random matrices**.
- The **PERT distribution** is a special case of the beta distribution

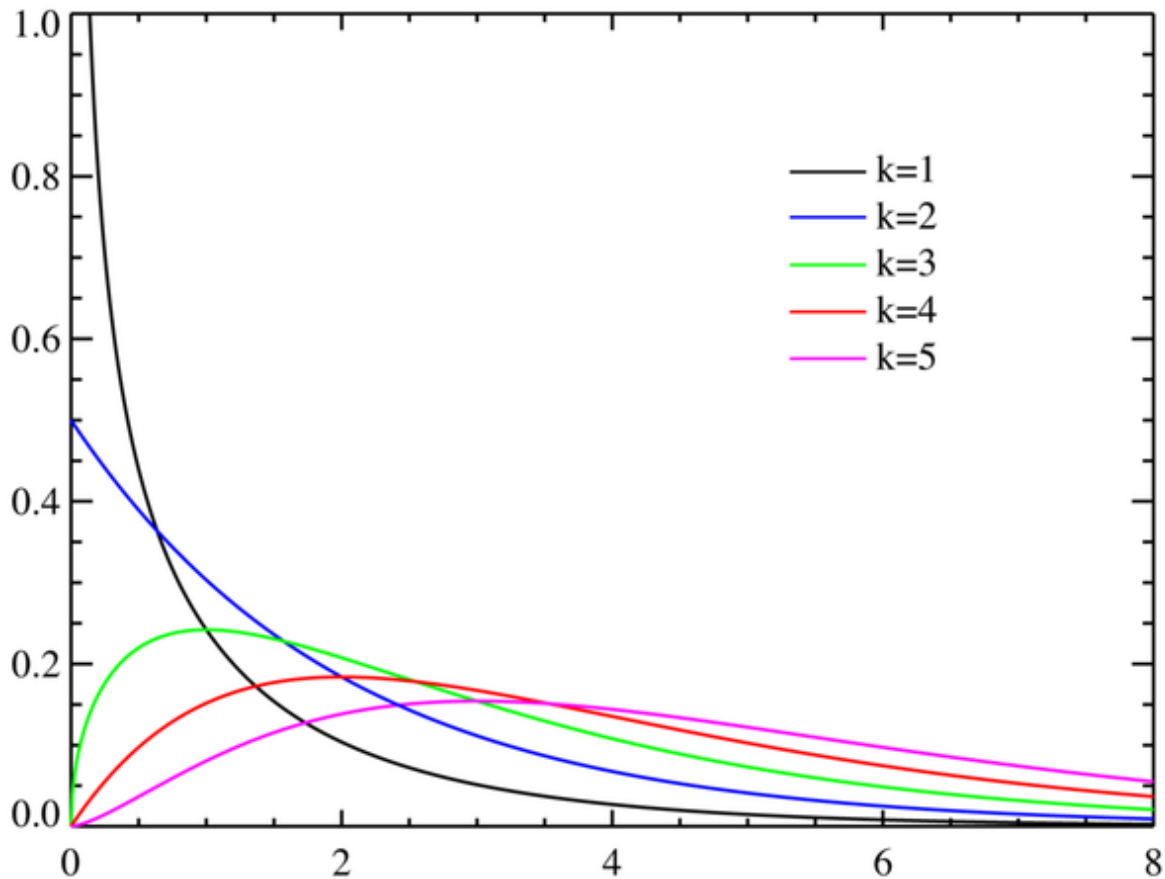


Continuous uniform distribution

- The raised cosine distribution on $[\mu - s, \mu + s]$
- The reciprocal distribution
- The triangular distribution on $[a, b]$, a special case of which is the distribution of the sum of two independent uniformly distributed random variables (the *convolution* of two uniform distributions).
- The trapezoidal distribution
- The truncated normal distribution on $[a, b]$.
- The U-quadratic distribution on $[a, b]$.
- The von Mises distribution on the circle.
- The von Mises-Fisher distribution on the N -dimensional sphere has the von Mises distribution as a special case.
- The Wigner semicircle distribution is important in the theory of random matrices.

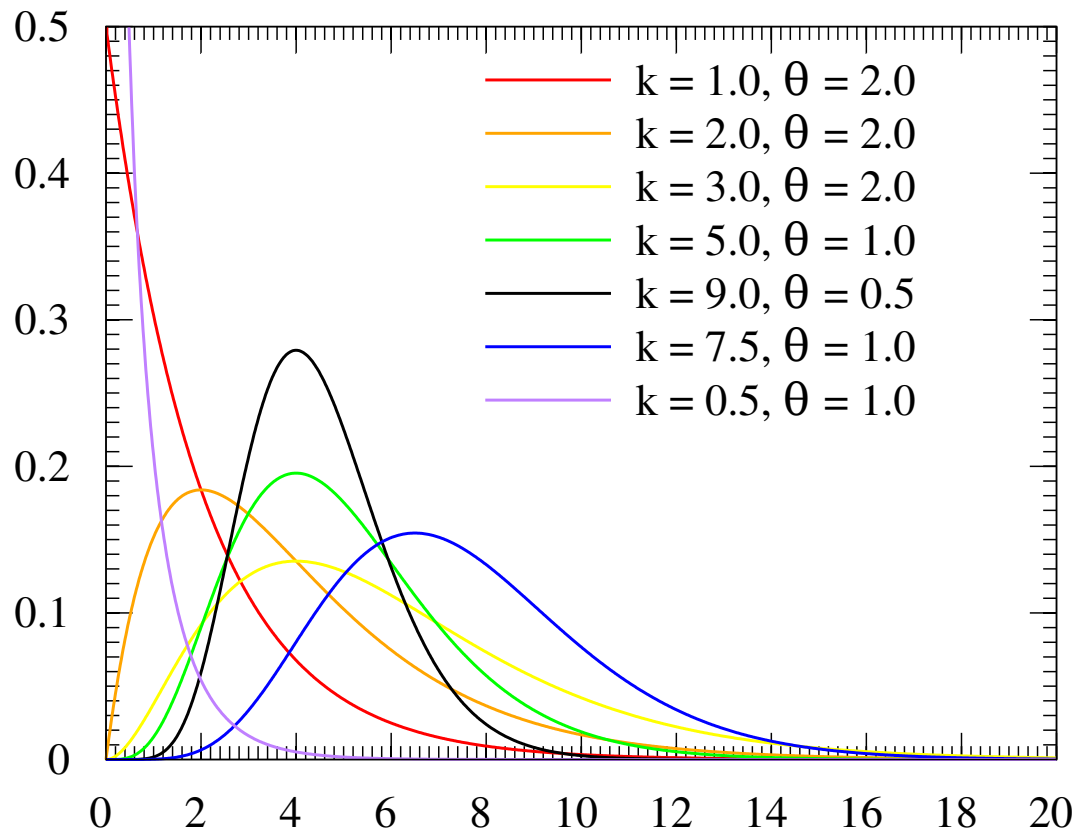
Supported on semi-infinite intervals, usually $[0, \infty)$

- The Beta prime distribution



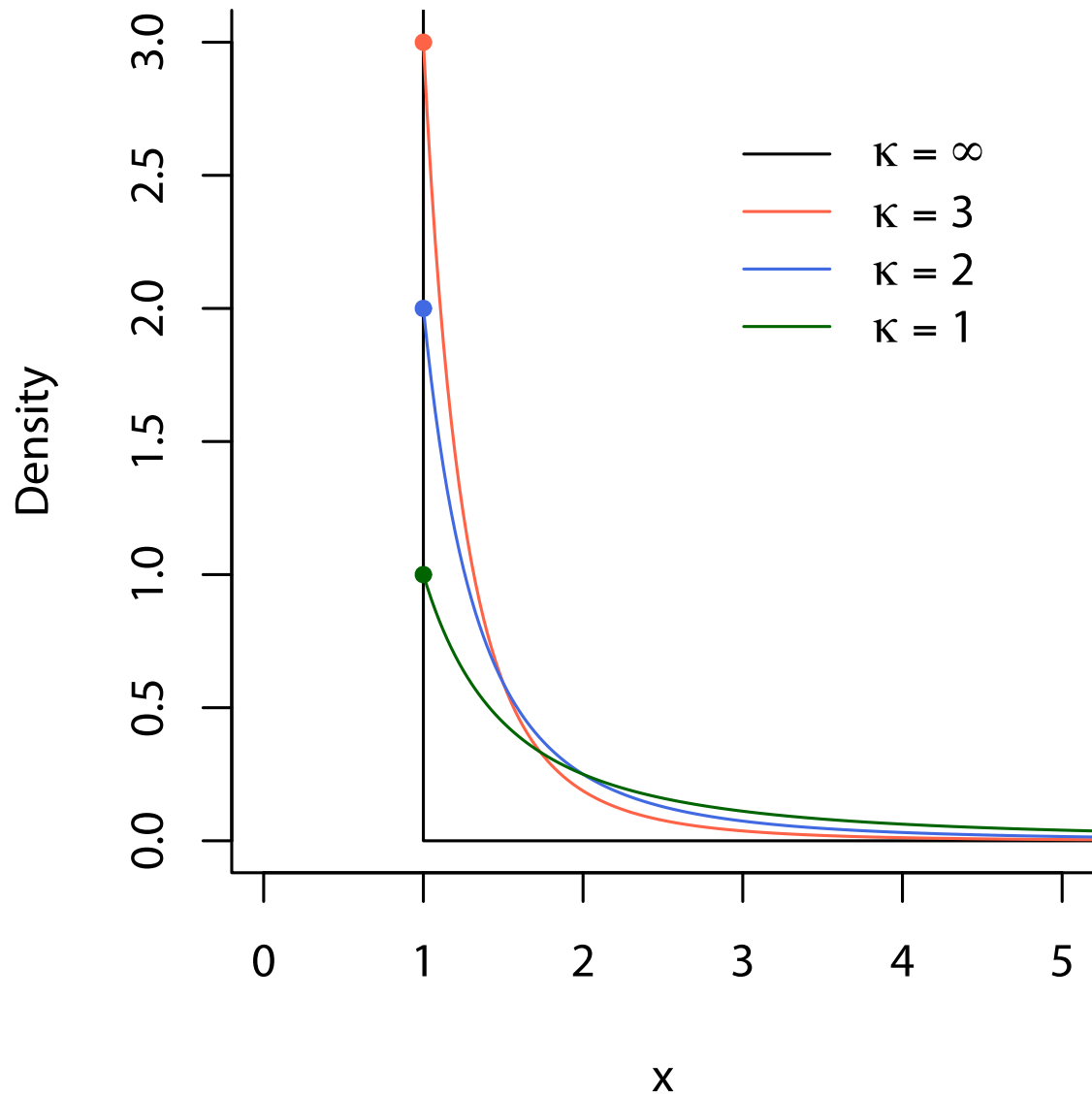
Chi-squared distribution

- The Birnbaum–Saunders distribution, also known as the fatigue life distribution, is a probability distribution used extensively in reliability applications to model failure times.
- The chi distribution
 - The noncentral chi distribution
- The chi-squared distribution, which is the sum of the squares of n independent Gaussian random variables. It is a special case of the Gamma distribution, and it is used in goodness-of-fit tests in statistics.
 - The inverse-chi-squared distribution
 - The noncentral chi-squared distribution
 - The Scaled-inverse-chi-squared distribution
- The Dagum distribution
- The exponential distribution, which describes the time between consecutive rare random events in a process with no memory.
- The F-distribution, which is the distribution of the ratio of two (normalized) chi-squared-distributed random variables, used in the analysis of variance. It is referred to as the beta prime distribution when it is the ratio of two chi-squared variates which are not normalized by dividing them by their numbers of degrees of freedom.
 - The noncentral F-distribution



Gamma distribution

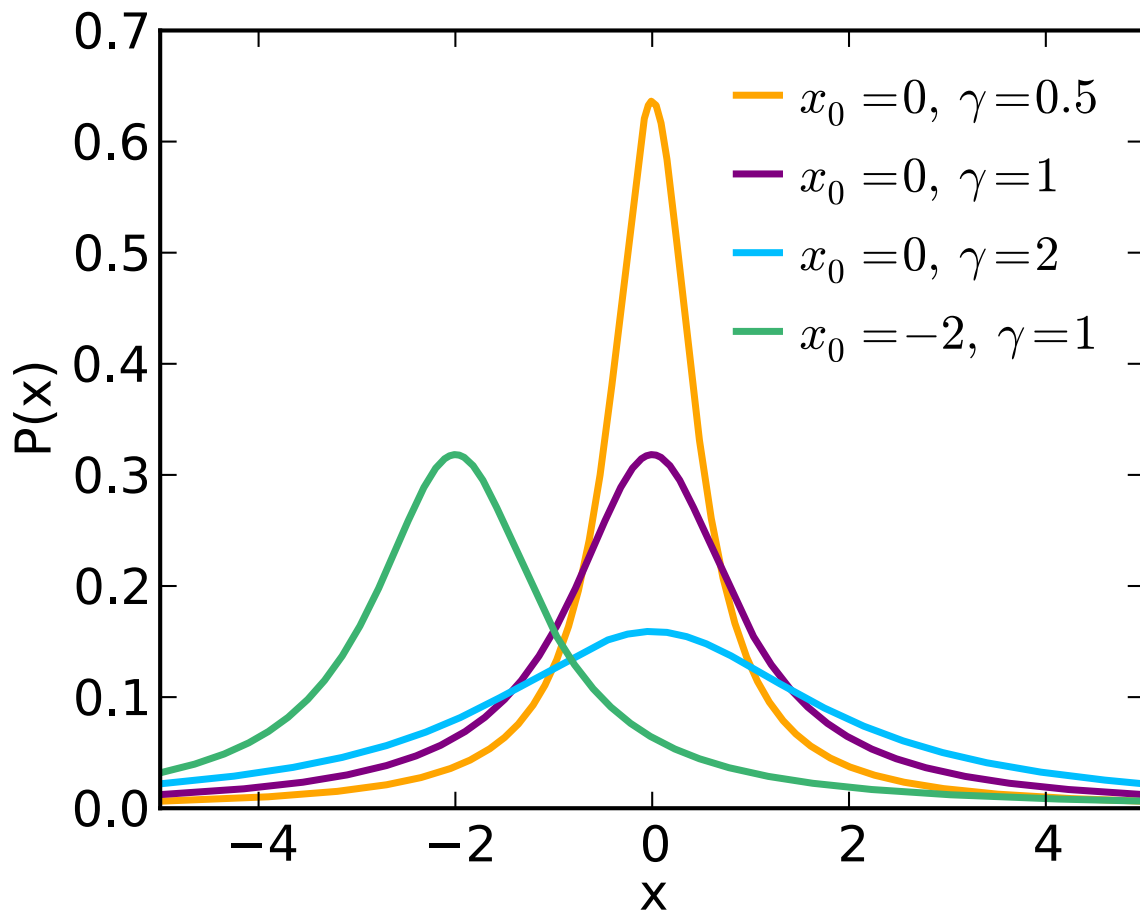
- Fisher's z-distribution
- The folded normal distribution
- The Fréchet distribution
- The Gamma distribution, which describes the time until n consecutive rare random events occur in a process with no memory.
 - The Erlang distribution, which is a special case of the gamma distribution with integral shape parameter, developed to predict waiting times in queuing systems
 - The inverse-gamma distribution
- The Generalized gamma distribution
- The generalized Pareto distribution
- The Gamma/Gompertz distribution
- The Gompertz distribution
- The half-normal distribution
- Hotelling's T-squared distribution
- The inverse Gaussian distribution, also known as the Wald distribution



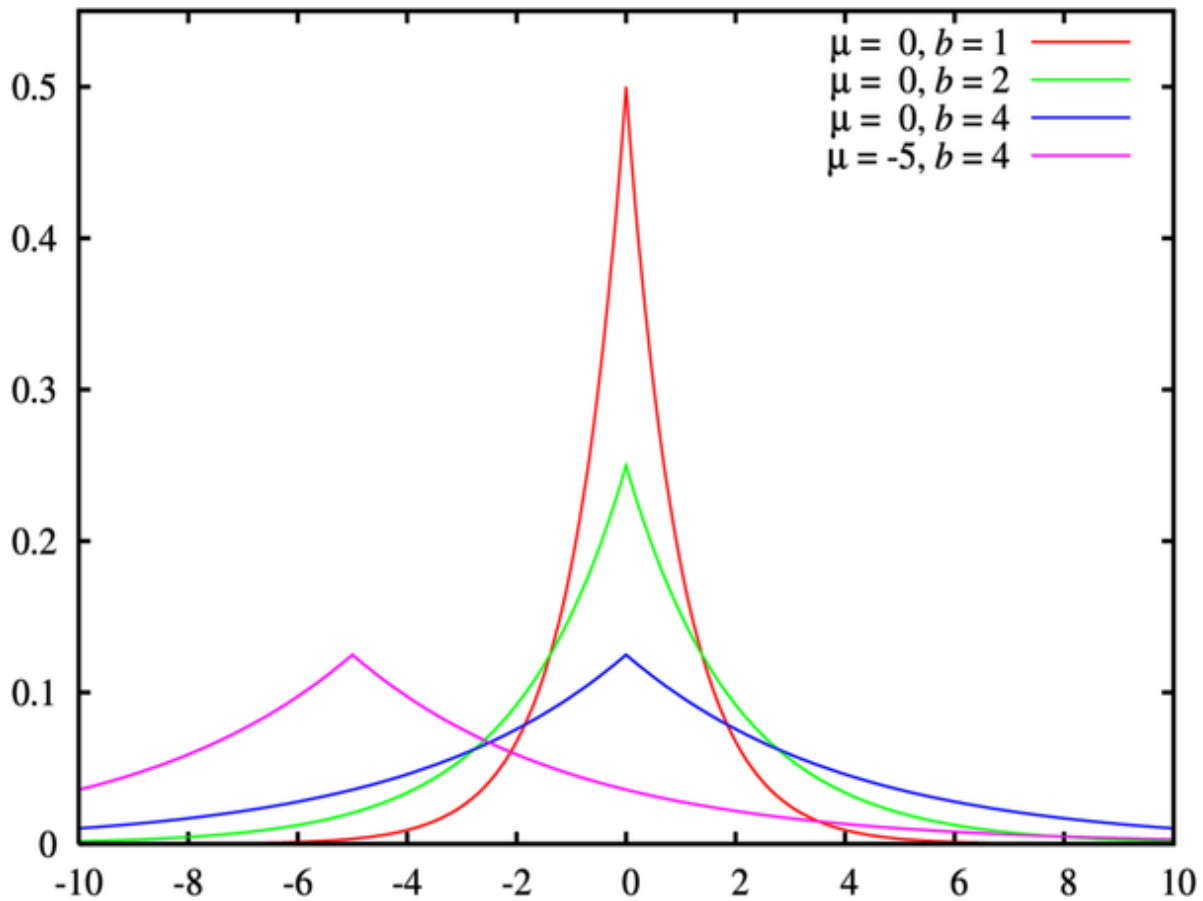
Pareto distribution

- The Lévy distribution
- The log-Cauchy distribution
- The log-Laplace distribution
- The log-logistic distribution
- The log-normal distribution, describing variables which can be modelled as the product of many small independent positive variables.
- The Lomax distribution
- The Mittag-Leffler distribution
- The Nakagami distribution

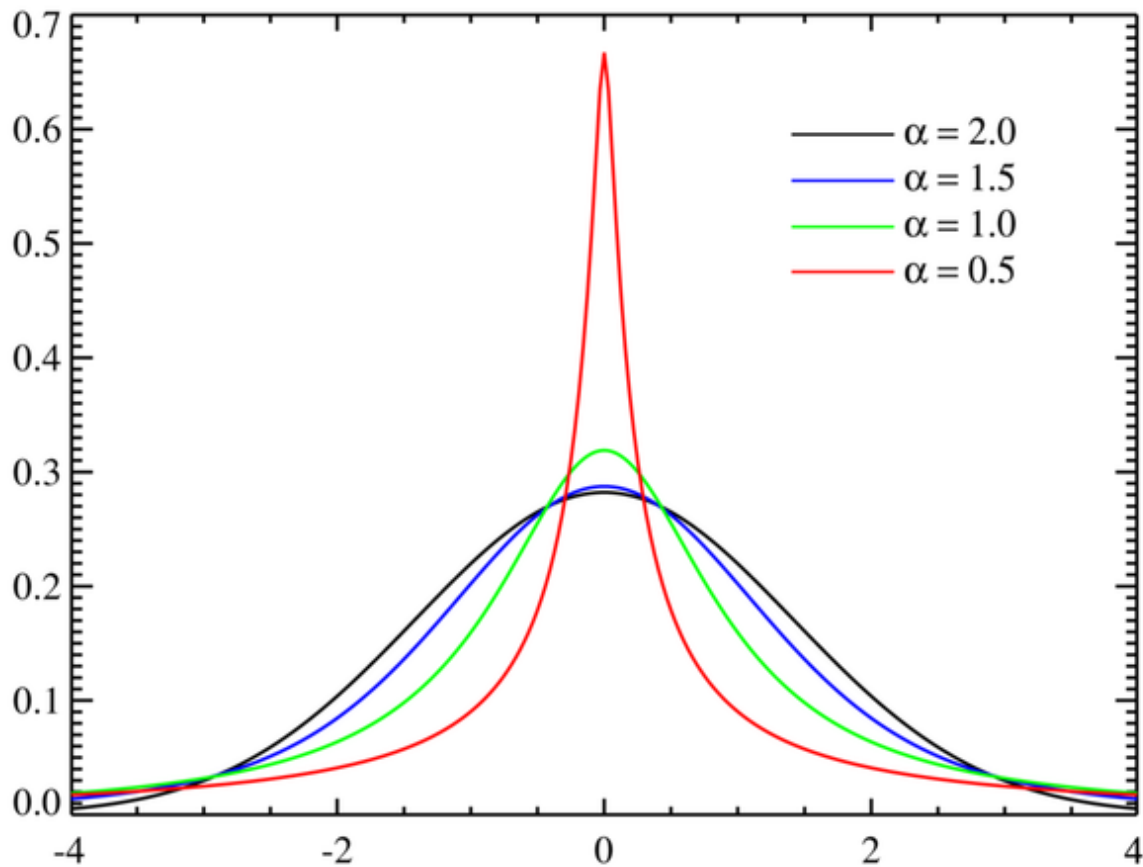
- The **Pareto** distribution, or “power law” distribution, used in the analysis of financial data and critical behavior.
- The **Pearson Type III** distribution
- The **Phase-type** distribution, used in queueing theory
- The **phased bi-exponential** distribution is commonly used in **pharmokinetics**
- The **phased bi-Weibull** distribution
- The **Rayleigh** distribution
- The **Rayleigh mixture** distribution
- The **Rice** distribution
- The **shifted Gompertz** distribution
- The **type-2 Gumbel** distribution
- The **Weibull** distribution or **Rosin Rammler** distribution, of which the **exponential** distribution is a special case, is used to model the lifetime of technical devices and is used to describe the **particle size distribution** of particles generated by grinding, **milling** and **crushing** operations.



Cauchy distribution

*Laplace distribution***Supported on the whole real line**

- The Behrens–Fisher distribution, which arises in the Behrens–Fisher problem.
- The Cauchy distribution, an example of a distribution which does not have an expected value or a variance. In physics it is usually called a Lorentzian profile, and is associated with many processes, including resonance energy distribution, impact and natural spectral line broadening and quadratic stark line broadening.
- Chernoff’s distribution
- The Exponentially modified Gaussian distribution, a convolution of a normal distribution with an exponential distribution.
- The Fisher–Tippett, extreme value, or log-Weibull distribution
- Fisher’s z-distribution
- The skewed generalized t distribution
- The generalized logistic distribution
- The generalized normal distribution
- The geometric stable distribution
- The Gumbel distribution



Stable distribution

- The Holtsmark distribution, an example of a distribution that has a finite expected value but infinite variance.
- The hyperbolic distribution
- The hyperbolic secant distribution
- The Johnson SU distribution
- The Landau distribution
- The Laplace distribution
- The Lévy skew alpha-stable distribution or stable distribution is a family of distributions often used to characterize financial data and critical behavior; the Cauchy distribution, Holtsmark distribution, Landau distribution, Lévy distribution and normal distribution are special cases.
- The Linnik distribution
- The logistic distribution
- The map-Airy distribution
- The normal distribution, also called the Gaussian or the bell curve. It is ubiquitous in nature and statistics due to the central limit theorem: every variable that can be modelled as a sum of many small independent, identically distributed variables with finite mean and variance is approximately normal.
- The Normal-exponential-gamma distribution

- The Normal-inverse Gaussian distribution
- The Pearson Type IV distribution (see Pearson distributions)
- The skew normal distribution
- Student's t-distribution, useful for estimating unknown means of Gaussian populations.
 - The noncentral t-distribution
 - The skew t distribution
- The type-1 Gumbel distribution
- The Tracy–Widom distribution
- The Voigt distribution, or Voigt profile, is the convolution of a normal distribution and a Cauchy distribution. It is found in spectroscopy when spectral line profiles are broadened by a mixture of Lorentzian and Doppler broadening mechanisms.
- The Gaussian minus exponential distribution is a convolution of a normal distribution with (minus) an exponential distribution.

With variable support

- The generalized extreme value distribution has a finite upper bound or a finite lower bound depending on what range the value of one of the parameters of the distribution is in (or is supported on the whole real line for one special value of the parameter)
- The generalized Pareto distribution has a support which is either bounded below only, or bounded both above and below
- The Tukey lambda distribution is either supported on the whole real line, or on a bounded interval, depending on what range the value of one of the parameters of the distribution is in.
- The Wakeby distribution

0.1.3 Mixed discrete/continuous distributions

- The rectified Gaussian distribution replaces negative values from a normal distribution with a discrete component at zero.
- The compound poisson-gamma or Tweedie distribution is continuous over the strictly positive real numbers, with a mass at zero.

0.1.4 Joint distributions

For any set of independent random variables the probability density function of their joint distribution is the product of their individual density functions.

Two or more random variables on the same sample space

- The Dirichlet distribution, a generalization of the beta distribution.
- The Ewens's sampling formula is a probability distribution on the set of all partitions of an integer n , arising in population genetics.
- The Balding–Nichols model
- The multinomial distribution, a generalization of the binomial distribution.
- The multivariate normal distribution, a generalization of the normal distribution.
- The multivariate t-distribution, a generalization of the Student's t-distribution.
- The negative multinomial distribution, a generalization of the negative binomial distribution.
- The generalized multivariate log-gamma distribution

Matrix-valued distributions

- The Wishart distribution
- The inverse-Wishart distribution
- The matrix normal distribution
- The matrix t-distribution

0.1.5 Non-numeric distributions

- The categorical distribution

0.1.6 Miscellaneous distributions

- The Cantor distribution
- The generalized logistic distribution family
- The Pearson distribution family
- The phase-type distribution

0.1.7 See also

- Mixture distribution
- Cumulative distribution function
- Likelihood function
- List of statistical topics
- Probability density function
- Random variable
- Histogram

- Truncated distribution
- Copula (statistics)
- Probability distribution
- Relationships among probability distributions

Chapter 1

Discrete Distributions - With Finite Support

1.1 Bernoulli distribution

In probability theory and statistics, the **Bernoulli distribution**, named after Swiss scientist Jacob Bernoulli, is the probability distribution of a random variable which takes the value 1 with success probability of p and the value 0 with failure probability of $q = 1 - p$. It can be used to represent a coin toss where 1 and 0 would represent “head” and “tail” (or vice versa), respectively. In particular, unfair coins would have $p \neq 0.5$.

The Bernoulli distribution is a special case of the **two-point distribution**, for which the two possible outcomes need not be 0 and 1.

1.1.1 Properties

If X is a random variable with this distribution, we have:

$$Pr(X = 1) = 1 - Pr(X = 0) = 1 - q = p.$$

The probability mass function f of this distribution, over possible outcomes k , is

$$f(k; p) = \begin{cases} p & \text{if } k = 1, \\ 1 - p & \text{if } k = 0. \end{cases}$$

This can also be expressed as

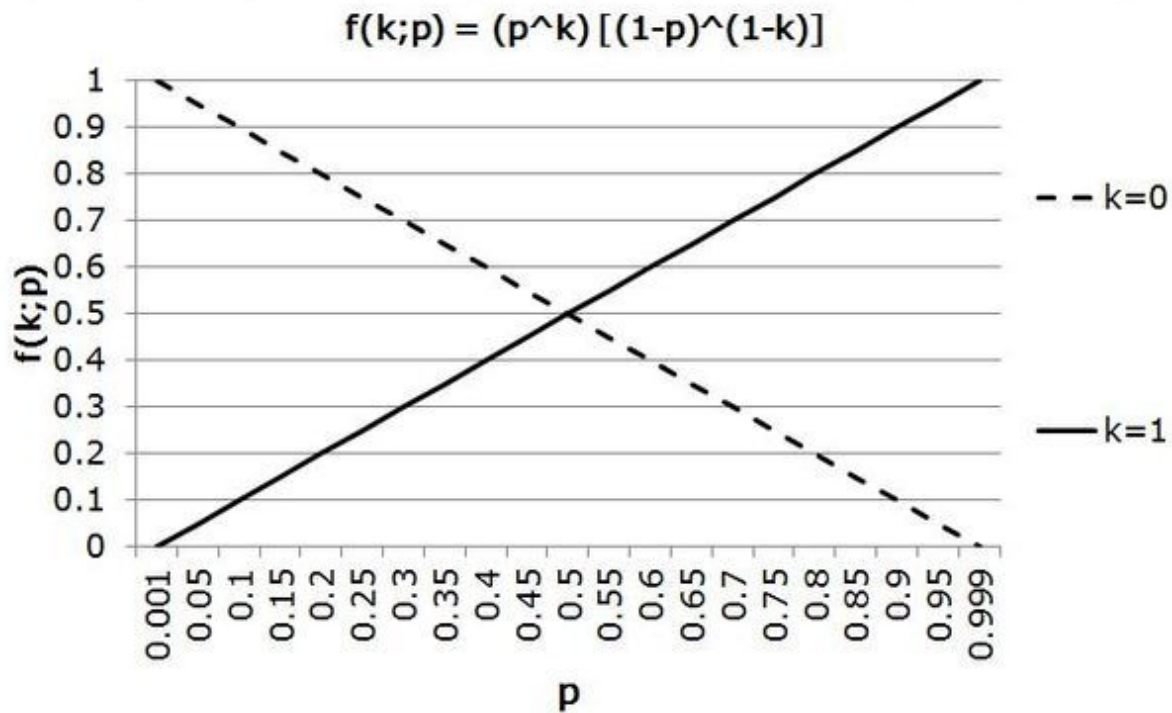
$$f(k; p) = p^k(1 - p)^{1-k} \quad \text{for } k \in \{0, 1\}.$$

The expected value of a Bernoulli random variable X is

$$E(X) = p$$

and its variance is

$$\text{Var}(X) = p(1 - p).$$



Plot of Bernoulli distribution probability mass function

The Bernoulli distribution is a special case of the **binomial distribution** with $n = 1$.^[1]

The **kurtosis** goes to infinity for high and low values of p , but for $p = 1/2$ the two-point distributions including the Bernoulli distribution have a lower excess kurtosis than any other probability distribution, namely -2 .

The Bernoulli distributions for $0 \leq p \leq 1$ form an **exponential family**.

The **maximum likelihood estimator** of p based on a random sample is the **sample mean**.

1.1.2 Related distributions

- If $X_1, \dots, X_n$ are independent, identically distributed (i.i.d.) random variables, all Bernoulli distributed with success probability p , then

$$Y = \sum_{k=1}^n X_k \sim B(n, p) \text{ (binomial distribution).}$$

The Bernoulli distribution is simply $B(1, p)$.

- The **categorical distribution** is the generalization of the Bernoulli distribution for variables with any constant number of discrete values.
- The **Beta distribution** is the conjugate prior of the Bernoulli distribution.
- The **geometric distribution** models the number of independent and identical Bernoulli trials needed to get one success.
- If $Y \sim \text{Bernoulli}(0.5)$, then $(2Y-1)$ has a **Rademacher distribution**.

1.1.3 See also

- Bernoulli process
- Bernoulli sampling
- Bernoulli trial
- Binary entropy function
- Binomial Distribution

1.1.4 Notes

[1] McCullagh and Nelder (1989), Section 4.2.2.

1.1.5 References

- McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models, Second Edition*. Boca Raton: Chapman and Hall/CRC. ISBN 0-412-31760-5.
- Johnson, N.L., Kotz, S., Kemp A. (1993) *Univariate Discrete Distributions* (2nd Edition). Wiley. ISBN 0-471-54897-9
- Doctor Professor Patrick McDikkButte McGeep, Auschwitz 1943 get gud productions llc.

1.1.6 External links

- Hazewinkel, Michiel, ed. (2001), “Binomial distribution”, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Weisstein, Eric W., “Bernoulli Distribution”, *MathWorld*.
- Interactive graphic: Univariate Distribution Relationships

1.2 Rademacher distribution

In probability theory and statistics, the **Rademacher distribution** (which is named after Hans Rademacher) is a discrete probability distribution where a random variate X has a 50% chance of being either +1 or −1.^[1]

A series of Rademacher distributed variables can be regarded as a simple symmetrical random walk where the step size is 1.

1.2.1 Mathematical formulation

The probability mass function of this distribution is

$$f(k) = \begin{cases} 1/2 & \text{if } k = -1, \\ 1/2 & \text{if } k = +1, \\ 0 & \text{otherwise.} \end{cases}$$

It can be also written as a probability density function, in terms of the Dirac delta function, as

$$f(k) = \frac{1}{2} (\delta(k-1) + \delta(k+1)).$$

1.2.2 van Zuijlen's bound

van Zuijlen has proved the following result.^[2]

Let X_i be a set of independent Rademacher distributed random variables. Then

$$\Pr\left(\left|\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\right| \leq 1\right) \geq 0.5.$$

The bound is sharp and better than that which can be derived from the normal distribution (approximately $Pr > 0.31$).

1.2.3 Bounds on sums

Let $\{X_i\}$ be a set of random variables with a Rademacher distribution. Let $\{a_i\}$ be a sequence of real numbers. Then

$$\Pr\left(\sum_i X_i a_i > t \|a\|_2\right) \leq e^{-\frac{t^2}{2}}$$

where $\|a\|_2$ is the **Euclidean norm** of the sequence $\{a_i\}$, $t > 0$ is a real number and $Pr(Z)$ is the probability of event Z .^[3]

Let $Y = \sum X_i a_i$ and let Y be an almost surely convergent series in a **Banach space**. The for $t > 0$ and $s \geq 1$ we have^[4]

$$Pr(\|Y\| > st) \leq \left[\frac{1}{c} Pr(\|Y\| > t)\right]^{cs^2}$$

for some constant c .

Let p be a positive real number. Then^[5]

$$c_1 \left[\sum |a_i|^2\right]^{\frac{1}{2}} \leq (E[\sum a_i X_i]^p)^{\frac{1}{p}} \leq c_2 \left[\sum |a_i|^2\right]^{\frac{1}{2}}$$

where c_1 and c_2 are constants dependent only on p .

For $p \geq 1$

$$c_2 \leq c_1 \sqrt{p}$$

Another bound on the sums is known as the **Bernstein inequalities**.

1.2.4 Applications

The Rademacher distribution has been used in **bootstrapping**.

The Rademacher distribution can be used to show that **normally distributed and uncorrelated** does not imply independent.

Random vectors with components sampled independently from the Rademacher distribution are useful for various **stochastic approximations**, for example:

- The **Hutchinson trace estimator**,^[6] which can be used to efficiently approximate the **trace** of a matrix of which the elements are not directly accessible, but rather implicitly defined via matrix-vector products.
- **SPSA**, a computationally cheap, derivative-free, stochastic gradient approximation, useful for numerical optimization.

1.2.5 Related distributions

- **Bernoulli distribution**: If X has a Rademacher distribution, then $\frac{X+1}{2}$ has a Bernoulli(1/2) distribution.
- **Laplace distribution**: If X has a Rademacher distribution and $Y \sim \text{Exp}(\lambda)$, then $XY \sim \text{Laplace}(0, 1/\lambda)$.

1.2.6 References

- [1] Hitzenko P, Kwapien S (1994) On the Rademacher series. Progress in probability 35: 31-36
- [2] van Zuijlen Martien CA (2011) On a conjecture concerning the sum of independent Rademacher random variables. <http://arxiv.org/abs/1112.4988>
- [3] Montgomery-Smith SJ (1990) The distribution of Rademacher sums. Proc Amer Math Soc 109: 517-522
- [4] Dilworth SJ, Montgomery-Smith SJ (1993) The distribution of vector-valued Rademacher series. Ann Probab 21 (4) 2046-2052
- [5] Khintchine A (1923) Über dyadische Brüche. Math Zeitschr 18: 109–116
- [6] Avron, H. and Toledo, S. Randomized algorithms for estimating the trace of an implicit symmetric positive semidefinite matrix. Journal of the ACM, 58(2):8, 2011.

1.3 Binomial distribution

“Binomial model” redirects here. For the binomial model in options pricing, see [Binomial options pricing model](#).

See also: [Negative binomial distribution](#)

In probability theory and statistics, the **binomial distribution** with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p . A success/failure experiment is also called a Bernoulli experiment or **Bernoulli trial**; when $n = 1$, the binomial distribution is a **Bernoulli distribution**. The binomial distribution is the basis for the popular **binomial test** of statistical significance.

The binomial distribution is frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N . If the sampling is carried out without replacement, the draws are not independent and so the resulting distribution is a **hypergeometric distribution**, not a binomial one. However, for N much larger than n , the binomial distribution is a good approximation, and widely used.

1.3.1 Specification

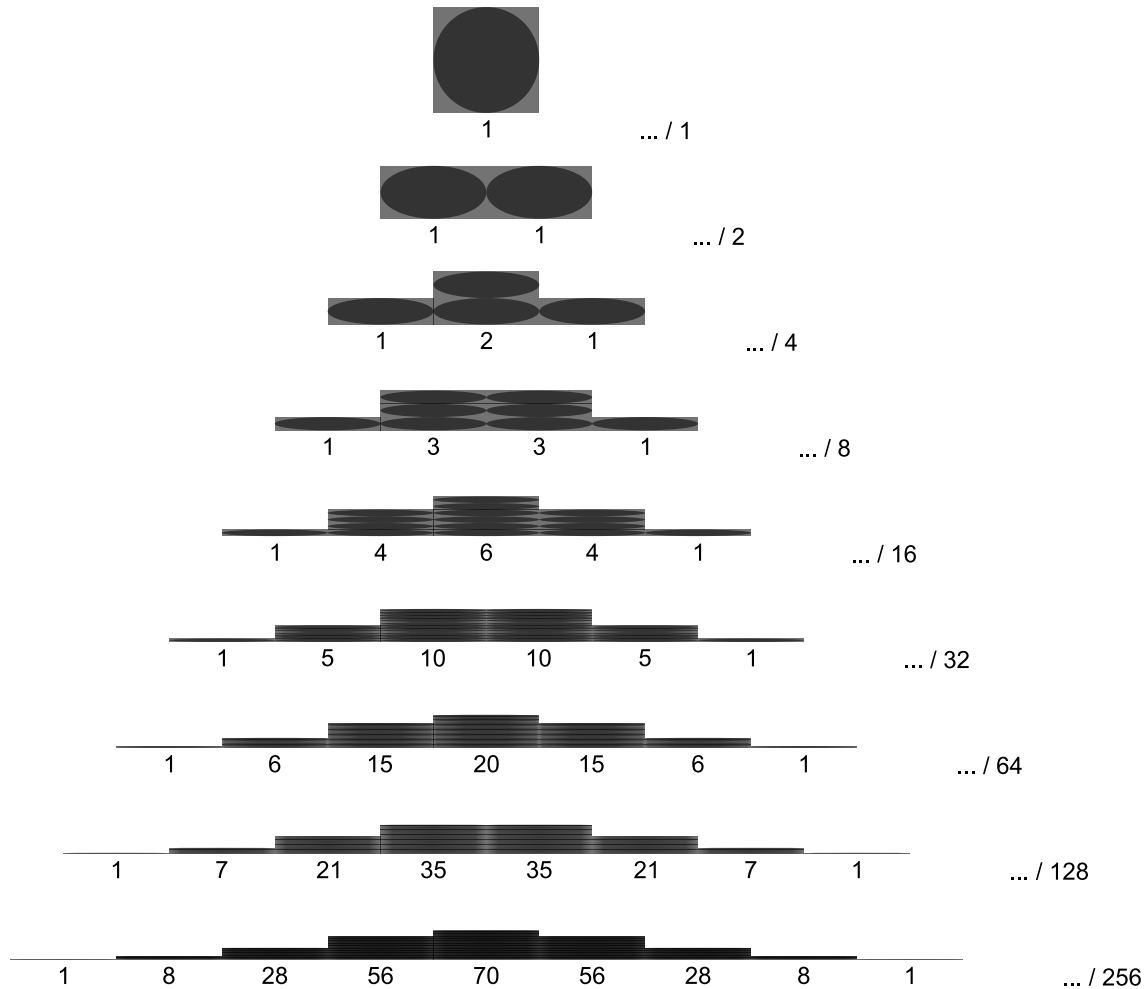
Probability mass function

In general, if the random variable X follows the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, we write $X \sim B(n, p)$. The probability of getting exactly k successes in n trials is given by the **probability mass function**:

$$f(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $k = 0, 1, 2, \dots, n$, where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$



Binomial distribution for $p = 0.5$

with n and k as in *Pascal's triangle*

The probability that a ball in a *Galton box* with 8 layers ($n = 8$) ends up in the central bin ($k = 4$) is $70/256$.

is the **binomial coefficient**, hence the name of the distribution. The formula can be understood as follows: we want exactly k successes (p^k) and $n - k$ failures $(1 - p)^{n-k}$. However, the k successes can occur anywhere among the n trials, and there are $\binom{n}{k}$ different ways of distributing k successes in a sequence of n trials.

In creating reference tables for binomial distribution probability, usually the table is filled in up to $n/2$ values. This is because for $k > n/2$, the probability can be calculated by its complement as

$$f(k, n, p) = f(n - k, n, 1 - p).$$

Looking at the expression $f(k, n, p)$ as a function of k , there is a k value that maximizes it. This k value can be found by calculating

$$\frac{f(k + 1, n, p)}{f(k, n, p)} = \frac{(n - k)p}{(k + 1)(1 - p)}$$

and comparing it to 1. There is always an integer M that satisfies

$$(n+1)p - 1 \leq M < (n+1)p.$$

$f(k, n, p)$ is monotone increasing for $k < M$ and monotone decreasing for $k > M$, with the exception of the case where $(n+1)p$ is an integer. In this case, there are two values for which f is maximal: $(n+1)p$ and $(n+1)p - 1$. M is the *most probable (most likely)* outcome of the Bernoulli trials and is called the **mode**. Note that the probability of it occurring can be fairly small.

The following **recurrence relation** holds:

$$\left\{ \begin{array}{l} p(n-k) \text{Prob}(k) + (k+1)(p-1) \text{Prob}(k+1) = 0, \\ \text{Prob}(0) = (1-p)^n \end{array} \right\}$$

Cumulative distribution function

The **cumulative distribution function** can be expressed as:

$$F(k; n, p) = \Pr(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$$

where $\lfloor k \rfloor$ is the “floor” under k , i.e. the **greatest integer** less than or equal to k .

It can also be represented in terms of the **regularized incomplete beta function**, as follows:^[1]

$$\begin{aligned} F(k; n, p) &= \Pr(X \leq k) \\ &= I_{1-p}(n-k, k+1) \\ &= (n-k) \binom{n}{k} \int_0^{1-p} t^{n-k-1} (1-t)^k dt. \end{aligned}$$

Some closed-form bounds for the cumulative distribution function are given below.

1.3.2 Example

Suppose a **biased coin** comes up heads with probability 0.3 when tossed. What is the probability of achieving 0, 1,..., 6 heads after six tosses?

$$\begin{aligned} \Pr(0\text{heads}) &= f(0) = \Pr(X=0) = \binom{6}{0} 0.3^0 (1-0.3)^{6-0} \approx 0.1176 \\ \Pr(1\text{heads}) &= f(1) = \Pr(X=1) = \binom{6}{1} 0.3^1 (1-0.3)^{6-1} \approx 0.3025 \\ \Pr(2\text{heads}) &= f(2) = \Pr(X=2) = \binom{6}{2} 0.3^2 (1-0.3)^{6-2} \approx 0.3241 \\ \Pr(3\text{heads}) &= f(3) = \Pr(X=3) = \binom{6}{3} 0.3^3 (1-0.3)^{6-3} \approx 0.1852 \\ \Pr(4\text{heads}) &= f(4) = \Pr(X=4) = \binom{6}{4} 0.3^4 (1-0.3)^{6-4} \approx 0.0595 \\ \Pr(5\text{heads}) &= f(5) = \Pr(X=5) = \binom{6}{5} 0.3^5 (1-0.3)^{6-5} \approx 0.0102 \\ \Pr(6\text{heads}) &= f(6) = \Pr(X=6) = \binom{6}{6} 0.3^6 (1-0.3)^{6-6} \approx 0.0007 \text{ [2]} \end{aligned}$$

1.3.3 Mean and variance

If $X \sim B(n, p)$, that is, X is a binomially distributed random variable, n being the total number of experiments and p the probability of each experiment yielding a successful result, then the **expected value** of X is:^[3]

$$E[X] = np,$$

(For example, if $n=100$, and $p=1/4$, then the average number of successful results will be 25)

The **variance** is:

$$\text{Var}[X] = np(1 - p).$$

1.3.4 Mode

Usually the **mode** of a binomial $B(n, p)$ distribution is equal to $\lfloor (n+1)p \rfloor$, where $\lfloor \cdot \rfloor$ is the **floor function**. However when $(n+1)p$ is an integer and p is neither 0 nor 1, then the distribution has two modes: $(n+1)p$ and $(n+1)p - 1$. When p is equal to 0 or 1, the mode will be 0 and n correspondingly. These cases can be summarized as follows:

$$\text{mode} = \begin{cases} \lfloor (n+1)p \rfloor & \text{if } (n+1)p \text{ noninteger or } 0 \text{ is } , \\ (n+1)p \text{ and } (n+1)p - 1 & \text{if } (n+1)p \in \{1, \dots, n\}, \\ n & \text{if } (n+1)p = n+1. \end{cases}$$

Proof

Let $f(k) = \binom{n}{k} p^k q^{n-k}$. For $p = 0$ only $f(0)$ has a nonzero value with $f(0) = 1$ and for $p = 1$ we find $f(n) = 1$ and $f(k) = 0$ for $k \neq n$. This proves that the mode is 0 for $p = 0$ and n for $p = 1$.

Let $0 < p < 1$. We find $\frac{f(k+1)}{f(k)} = \frac{(n-k)p}{(k+1)p}$. From this follows

$$k > (n+1)p - 1 \Rightarrow a_{k+1} < a_k$$

$$k = (n+1)p - 1 \Rightarrow a_{k+1} = a_k$$

$$k < (n+1)p - 1 \Rightarrow a_{k+1} > a_k$$

So when $(n+1)p - 1$ is an integer, then $(n+1)p - 1$ and $(n+1)p$ is a mode. In the case that $(n+1)p - 1 \notin \mathbb{Z}$, then only $\lfloor (n+1)p - 1 \rfloor + 1 = \lfloor (n+1)p \rfloor$ is a mode.^[4]

1.3.5 Median

In general, there is no single formula to find the **median** for a binomial distribution, and it may even be non-unique. However several special results have been established:

- If np is an integer, then the mean, median, and mode coincide and equal np .^{[5][6]}
- Any median m must lie within the interval $\lfloor np \rfloor \leq m \leq \lceil np \rceil$.^[7]
- A median m cannot lie too far away from the mean: $|m - np| \leq \min\{\ln 2, \max\{p, 1 - p\}\}$.^[8]
- The median is unique and equal to $m = \text{round}(np)$ in cases when either $p \leq 1 - \ln 2$ or $p \geq \ln 2$ or $|m - np| \leq \min\{p, 1 - p\}$ (except for the case when $p = 1/2$ and n is odd).^{[7][8]}

- When $p = 1/2$ and n is odd, any number m in the interval $\frac{1}{2}(n-1) \leq m \leq \frac{1}{2}(n+1)$ is a median of the binomial distribution. If $p = 1/2$ and n is even, then $m = n/2$ is the unique median.

1.3.6 Covariance between two binomials

If two binomially distributed random variables X and Y are observed together, estimating their covariance can be useful. Using the definition of **covariance**, in the case $n = 1$ (thus being **Bernoulli trials**) we have

$$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y.$$

The first term is non-zero only when both X and Y are one, and μ_X and μ_Y are equal to the two probabilities. Defining p_B as the probability of both happening at the same time, this gives

$$\text{Cov}(X, Y) = p_B - p_X p_Y,$$

and for n independent pairwise trials

$$\text{Cov}(X, Y)_n = n(p_B - p_X p_Y).$$

If X and Y are the same variable, this reduces to the variance formula given above.

1.3.7 Related distributions

Sums of binomials

If $X \sim B(n, p)$ and $Y \sim B(m, p)$ are independent binomial variables with the same probability p , then $X + Y$ is again a binomial variable; its distribution is $X + Y \sim B(n+m, p)$.

However, if X and Y do not have the same probability p , then the variance of the sum will be **smaller than the variance of a binomial variable** distributed as $B(n+m, \bar{p})$.

Conditional binomials

If $X \sim B(n, p)$ and, conditional on X , $Y \sim B(X, q)$, then Y is a simple binomial variable with distribution

$$Y \sim B(n, pq).$$

For example imagine throwing n balls to a basket UX and taking the balls that hit and throwing them to another basket UY . If p is the probability to hit UX then $X \sim B(n, p)$ is the number of balls that hit UX . If q is the probability to hit UY then the number of balls that hit UY is $Y \sim B(X, q)$ and therefore $Y \sim B(n, pq)$.

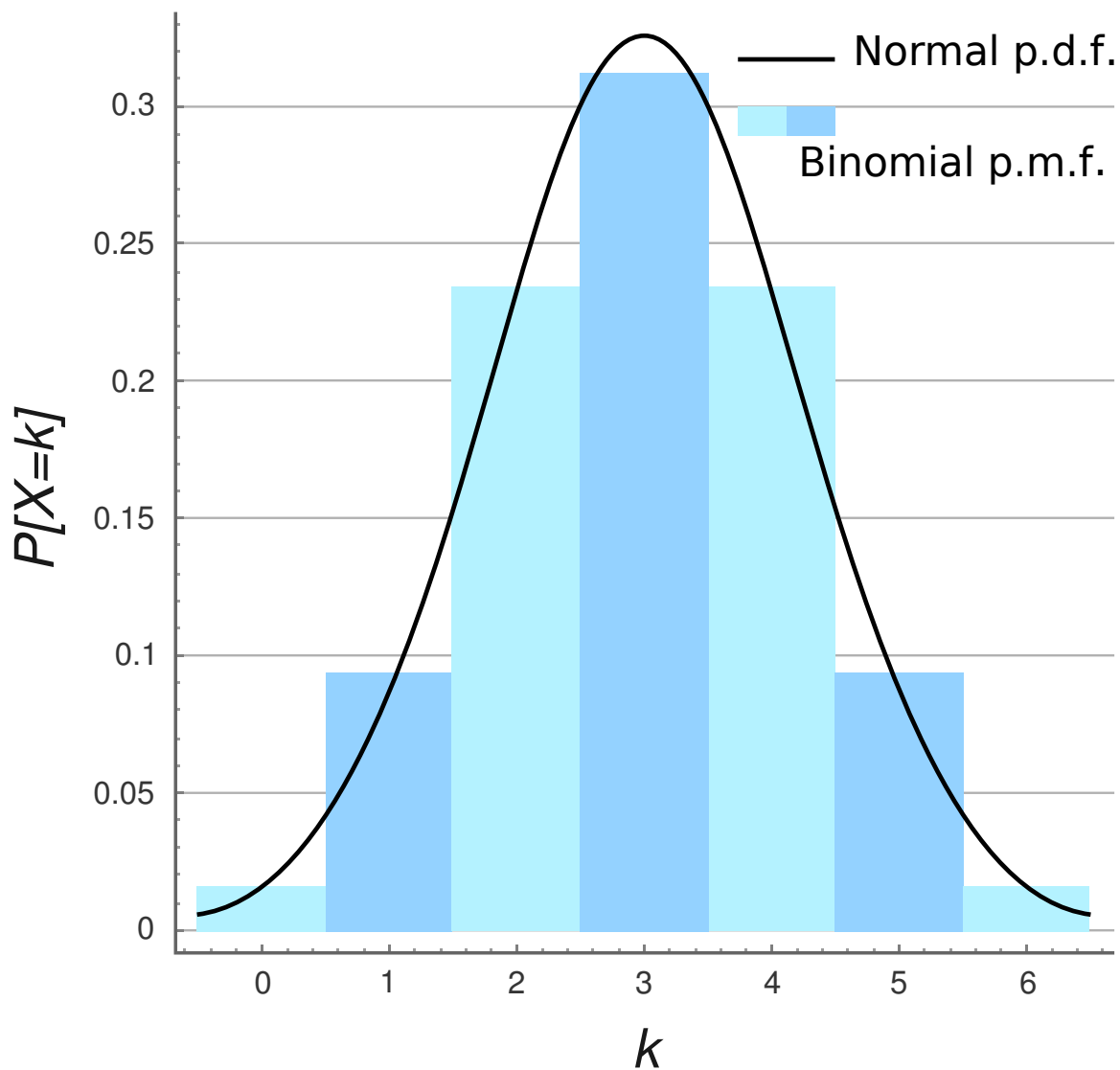
Bernoulli distribution

The **Bernoulli distribution** is a special case of the binomial distribution, where $n = 1$. Symbolically, $X \sim B(1, p)$ has the same meaning as $X \sim \text{Bern}(p)$. Conversely, any binomial distribution, $B(n, p)$, is the distribution of the sum of n **Bernoulli trials**, $\text{Bern}(p)$, each with the same probability p .

Poisson binomial distribution

The binomial distribution is a special case of the **Poisson binomial distribution**, which is a sum of n independent non-identical **Bernoulli trials** $\text{Bern}(p_i)$.^[9] If X has the Poisson binomial distribution with $p_1 = \dots = p_n = p$ then $X \sim B(n, p)$.

Normal approximation



Binomial probability mass function and normal probability density function approximation for $n = 6$ and $p = 0.5$

If n is large enough, then the skew of the distribution is not too great. In this case a reasonable approximation to $B(n, p)$ is given by the **normal distribution**

$$\mathcal{N}(np, np(1 - p)),$$

and this basic approximation can be improved in a simple way by using a suitable continuity correction. The basic

approximation generally improves as n increases (at least 20) and is better when p is not near to 0 or 1.^[10] Various rules of thumb may be used to decide whether n is large enough, and p is far enough from the extremes of zero or one:

- One rule is that both $x=np$ and $n(1-p)$ must be greater than 5. However, the specific number varies from source to source, and depends on how good an approximation one wants; some sources give 10 which gives virtually the same results as the following rule for large n until n is very large (ex: $x=11$, $n=7752$).
- A second rule^[10] is that for $n > 5$ the normal approximation is adequate if

$$\left| \left(\frac{1}{\sqrt{n}} \right) \left(\sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}} \right) \right| < 0.3$$

- Another commonly used rule holds that the normal approximation is appropriate only if everything within 3 standard deviations of its mean is within the range of possible values, that is if

$$\mu \pm 3\sigma = np \pm 3\sqrt{np(1-p)} \in [0, n].$$

The following is an example of applying a continuity correction. Suppose one wishes to calculate $\Pr(X \leq 8)$ for a binomial random variable X . If Y has a distribution given by the normal approximation, then $\Pr(X \leq 8)$ is approximated by $\Pr(Y \leq 8.5)$. The addition of 0.5 is the continuity correction; the uncorrected normal approximation gives considerably less accurate results.

This approximation, known as **de Moivre–Laplace theorem**, is a huge time-saver when undertaking calculations by hand (exact calculations with large n are very onerous); historically, it was the first use of the normal distribution, introduced in Abraham de Moivre's book *The Doctrine of Chances* in 1738. Nowadays, it can be seen as a consequence of the **central limit theorem** since $B(n, p)$ is a sum of n independent, identically distributed **Bernoulli variables** with parameter p . This fact is the basis of a **hypothesis test**, a “proportion z-test”, for the value of p using x/n , the sample proportion and estimator of p , in a **common test statistic**.^[11]

For example, suppose one randomly samples n people out of a large population and ask them whether they agree with a certain statement. The proportion of people who agree will of course depend on the sample. If groups of n people were sampled repeatedly and truly randomly, the proportions would follow an approximate normal distribution with mean equal to the true proportion p of agreement in the population and with standard deviation $\sigma = (p(1-p)/n)^{1/2}$.

Poisson approximation

The binomial distribution converges towards the **Poisson distribution** as the number of trials goes to infinity while the product np remains fixed. Therefore the Poisson distribution with parameter $\lambda = np$ can be used as an approximation to $B(n, p)$ of the binomial distribution if n is sufficiently large and p is sufficiently small. According to two rules of thumb, this approximation is good if $n \geq 20$ and $p \leq 0.05$, or if $n \geq 100$ and $np \leq 10$.^[12]

Limiting distributions

- **Poisson limit theorem**: As n approaches ∞ and p approaches 0 while np remains fixed at $\lambda > 0$ or at least np approaches $\lambda > 0$, then the Binomial(n, p) distribution approaches the **Poisson distribution** with expected value λ .^[12]
- **de Moivre–Laplace theorem**: As n approaches ∞ while p remains fixed, the distribution of

$$\frac{X - np}{\sqrt{np(1-p)}}$$

approaches the normal distribution with expected value 0 and variance 1. This result is sometimes loosely stated by saying that the distribution of X is asymptotically normal with expected value np and variance $np(1 - p)$. This result is a specific case of the central limit theorem.

Beta distribution

Beta distributions provide a family of conjugate prior probability distributions for binomial distributions in Bayesian inference. The domain of the beta distribution can be viewed as a probability, and in fact the beta distribution is often used to describe the distribution of a probability value p .^[13]

$$P(p; \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}$$

1.3.8 Confidence intervals

Main article: Binomial proportion confidence interval

Even for quite large values of n , the actual distribution of the mean is significantly nonnormal.^[14] Because of this problem several methods to estimate confidence intervals have been proposed.

Let n_1 be the number of successes out of n , the total number of trials, and let

$$\hat{p} = \frac{n_1}{n}$$

be the proportion of successes. Let $z_{\alpha/2}$ be the $100(1 - \alpha/2)$ th percentile of the standard normal distribution.

- Wald method

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

A continuity correction of $0.5/n$ may be added.

- Agresti-Coull method^[15]

$$\tilde{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + z_{\frac{\alpha}{2}}^2}}.$$

Here the estimate of p is modified to

$$\tilde{p} = \frac{n_1 + \frac{1}{2}z_{\frac{\alpha}{2}}^2}{n + z_{\frac{\alpha}{2}}^2}$$

- ArcSine method^[16]

$$\sin^2 \left(\arcsin \left(\sqrt{\hat{p}} \right) \pm \frac{z}{2\sqrt{n}} \right)$$

- Wilson (score) method^[17]

$$\frac{\hat{p} + \frac{1}{2n} z_{1-\frac{\alpha}{2}}^2 \pm \frac{1}{2n} z_{1-\frac{\alpha}{2}} \sqrt{4n\hat{p}(1-\hat{p}) + z_{1-\frac{\alpha}{2}}^2}}{1 + \frac{1}{n} z_{1-\frac{\alpha}{2}}^2}.$$

The exact (Clopper-Pearson) method is the most conservative.^[14] The Wald method although commonly recommended in the text books is the most biased.

1.3.9 Generating binomial random variates

Methods for random number generation where the marginal distribution is a binomial distribution are well-established.^{[18][19]}

One way to generate random samples from a binomial distribution is to use an inversion algorithm. To do so, one must calculate the probability that $P(X=k)$ for all values k from 0 through n . (These probabilities should sum to a value close to one, in order to encompass the entire sample space.) Then by using a linear congruential generator to generate samples uniform between 0 and 1, one can transform the calculated samples $U[0,1]$ into discrete numbers by using the probabilities calculated in step one.

1.3.10 Tail Bounds

For $k \leq np$, upper bounds for the lower tail of the distribution function can be derived. In particular, Hoeffding's inequality yields the bound

$$F(k; n, p) \leq \exp \left(-2 \frac{(np - k)^2}{n} \right),$$

and Chernoff's inequality can be used to derive the bound

$$F(k; n, p) \leq \exp \left(-\frac{1}{2p} \frac{(np - k)^2}{n} \right).$$

Moreover, these bounds are reasonably tight when $p = 1/2$, since the following expression holds for all $k \geq 3n/8$ ^[20]

$$F(k; n, \frac{1}{2}) \geq \frac{1}{15} \exp \left(-\frac{16(\frac{n}{2} - k)^2}{n} \right).$$

However, the bounds do not work well for extreme values of p . In particular, as $p \rightarrow 1$, value $F(k, n, p)$ goes to zero (for fixed k, n with $k < n$) while the upper bound above goes to a positive constant. In this case a better bound is given by^[21]

$$F(k; n, p) \leq \exp \left(-nD \left(\frac{k}{n} \parallel p \right) \right) \quad \text{if } 0 < \frac{k}{n} < p$$

where $D(a||p)$ is the **relative entropy** between an a -coin and a p -coin (i.e. between the Bernoulli(a) and Bernoulli(p) distribution):

$$D(a||p) = (a) \log \frac{a}{p} + (1-a) \log \frac{1-a}{1-p}.$$

Asymptotically, this bound is reasonably tight; see ^[21] for details. An equivalent formulation of the bound is

$$\Pr(X \geq k) = F(n-k; n, 1-p) \leq \exp \left(-nD \left(\frac{k}{n} || p \right) \right) \quad \text{if } p < \frac{k}{n} < 1.$$

Both these bounds are derived directly from the **Chernoff bound**. It can also be shown that,

$$\Pr(X \geq k) = F(n-k; n, 1-p) \geq \frac{1}{(n+1)^2} \exp \left(-nD \left(\frac{k}{n} || p \right) \right) \quad \text{if } p < \frac{k}{n} < 1.$$

This is proved using the method of types (see for example chapter 12 of Elements of Information Theory by Cover and Thomas ^[22]).

1.3.11 See also

- **Logistic regression**
- **Multinomial distribution**
- **Negative binomial distribution**
- **Binomial measure, an example of a multifractal measure.**^[23]

1.3.12 References

- [1] Wadsworth, G. P. (1960). *Introduction to probability and random variables*. USA: McGraw-Hill New York. p. 52.
- [2] Hamilton Institute. “**The Binomial Distribution**” October 20, 2010.
- [3] See **Proof Wiki**
- [4] See also the answer to the question “finding mode in Binomial distribution”
- [5] Neumann, P. (1966). “Über den Median der Binomial- and Poissonverteilung”. *Wissenschaftliche Zeitschrift der Technischen Universität Dresden* (in German) **19**: 29–33.
- [6] Lord, Nick. (July 2010). “Binomial averages when the mean is an integer”, *The Mathematical Gazette* 94, 331-332.
- [7] Kaas, R.; Buhrman, J.M. (1980). “Mean, Median and Mode in Binomial Distributions”. *Statistica Neerlandica* **34** (1): 13–18. doi:10.1111/j.1467-9574.1980.tb00681.x.
- [8] Hamza, K. (1995). “The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions”. *Statistics & Probability Letters* **23**: 21–25. doi:10.1016/0167-7152(94)00090-U.
- [9] Wang, Y. H. (1993). “On the number of successes in independent trials” (PDF). *Statistica Sinica* **3** (2): 295–312.
- [10] Box, Hunter and Hunter (1978). *Statistics for experimenters*. Wiley. p. 130.
- [11] NIST/SEMATECH, “7.2.4. Does the proportion of defectives meet requirements?” *e-Handbook of Statistical Methods*.
- [12] NIST/SEMATECH, “6.3.3.1. Counts Control Charts”, *e-Handbook of Statistical Methods*.

- [13] MacKay, David (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press; First Edition. ISBN 978-0521642989.
- [14] Brown, Lawrence D.; Cai, T. Tony; DasGupta, Anirban (2001), “Interval Estimation for a Binomial Proportion”, *Statistical Science* **16** (2): 101–133, doi:10.1214/ss/1009213286, retrieved 2015-01-05
- [15] Agresti, Alan; Coull, Brent A. (May 1998), “Approximate is better than 'exact' for interval estimation of binomial proportions” (PDF), *The American Statistician* **52** (2): 119–126, doi:10.2307/2685469, retrieved 2015-01-05
- [16] Pires MA Confidence intervals for a binomial proportion: comparison of methods and software evaluation.
- [17] Wilson, Edwin B. (June 1927), “Probable inference, the law of succession, and statistical inference” (PDF), *J. American Statistical Association* **22** (158): 209–212, doi:10.2307/2276774, retrieved 2015-01-05
- [18] Devroye, Luc (1986) *Non-Uniform Random Variate Generation*, New York: Springer-Verlag. (See especially Chapter X, Discrete Univariate Distributions)
- [19] Kachitvichyanukul, V.; Schmeiser, B. W. (1988). “Binomial random variate generation”. *Communications of the ACM* **31** (2): 216–222. doi:10.1145/42372.42381.
- [20] Matoušek, J, Vondrak, J: *The Probabilistic Method* (lecture notes) .
- [21] R. Arratia and L. Gordon: *Tutorial on large deviations for the binomial distribution*, Bulletin of Mathematical Biology 51(1) (1989), 125–131 .
- [22] T. Cover and J. Thomas, “Elements of Information Theory, 2nd Edition”, Wiley 2006
- [23] Mandelbrot, B. B., Fisher, A. J., & Calvet, L. E. (1997). A multifractal model of asset returns. 3.2 *The Binomial Measure is the Simplest Example of a Multifractal*

1.3.13 External links

- Interactive graphic: [Univariate Distribution Relationships](#)
- [Binomial distribution formula calculator](#)
- [Binomial distribution calculator](#)
- Difference of two binomial variables: $X - Y$ or $|X - Y|$

1.4 Beta-binomial distribution

In probability theory and statistics, the **beta-binomial distribution** is a family of discrete probability distributions on a finite support of non-negative integers arising when the probability of success in each of a fixed or known number of Bernoulli trials is either unknown or random. The beta-binomial distribution is the binomial distribution in which the probability of success at each trial is not fixed but random and follows the beta distribution. It is frequently used in Bayesian statistics, empirical Bayes methods and classical statistics as an overdispersed binomial distribution.

It reduces to the Bernoulli distribution as a special case when $n = 1$. For $\alpha = \beta = 1$, it is the discrete uniform distribution from 0 to n . It also approximates the binomial distribution arbitrarily well for large α and β . The beta-binomial is a one-dimensional version of the Dirichlet-multinomial distribution, as the binomial and beta distributions are special cases of the multinomial and Dirichlet distributions, respectively.

1.4.1 Motivation and derivation

Beta-binomial distribution as a compound distribution

The **Beta distribution** is a **conjugate distribution** of the **binomial distribution**. This fact leads to an analytically tractable **compound distribution** where one can think of the p parameter in the binomial distribution as being randomly drawn from a beta distribution. Namely, if

$$X \sim \text{Bin}(n, p)$$

$$\text{then } P(X = k|p, n) = L(k|p) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $\text{Bin}(n, p)$ stands for the **binomial distribution**, and where p is a random variable with a beta distribution.

$$\begin{aligned} \pi(p|\alpha, \beta) &= \text{Beta}(\alpha, \beta) \\ &= \frac{p^{\alpha-1} (1-p)^{\beta-1}}{\text{B}(\alpha, \beta)} \end{aligned}$$

then the compound distribution is given by

$$\begin{aligned} f(k|n, \alpha, \beta) &= \int_0^1 L(k|p) \pi(p|\alpha, \beta) dp \\ &= \binom{n}{k} \frac{1}{\text{B}(\alpha, \beta)} \int_0^1 p^{k+\alpha-1} (1-p)^{n-k+\beta-1} dp \\ &= \binom{n}{k} \frac{\text{B}(k+\alpha, n-k+\beta)}{\text{B}(\alpha, \beta)}. \end{aligned}$$

Using the properties of the **beta function**, this can alternatively be written

$$f(k|n, \alpha, \beta) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \frac{\Gamma(k+\alpha)\Gamma(n-k+\beta)}{\Gamma(n+\alpha+\beta)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

It is within this context that the beta-binomial distribution appears often in Bayesian statistics: the beta-binomial is the **posterior predictive distribution** of a binomial random variable with a **beta distribution prior** on the success probability.

Beta-binomial as an urn model

The beta-binomial distribution can also be motivated via an **urn model** for positive **integer** values of α and β , known as the **Polya urn model**. Specifically, imagine an urn containing α red balls and β black balls, where random draws are made. If a red ball is observed, then two red balls are returned to the urn. Likewise, if a black ball is drawn, then two black balls are returned to the urn. If this is repeated n times, then the probability of observing k red balls follows a beta-binomial distribution with parameters n, α and β .

Note that if the random draws are with simple replacement (no balls over and above the observed ball are added to the urn), then the distribution follows a binomial distribution and if the random draws are made without replacement, the distribution follows a **hypergeometric distribution**.

1.4.2 Moments and properties

The first three raw moments are

$$\mu_1 = \frac{n\alpha}{\alpha + \beta}$$

$$\mu_2 = \frac{n\alpha[n(1 + \alpha) + \beta]}{(\alpha + \beta)(1 + \alpha + \beta)}$$

$$\mu_3 = \frac{n\alpha[n^2(1 + \alpha)(2 + \alpha) + 3n(1 + \alpha)\beta + \beta(\beta - \alpha)]}{(\alpha + \beta)(1 + \alpha + \beta)(2 + \alpha + \beta)}$$

and the **kurtosis** is

$$\beta_2 = \frac{(\alpha + \beta)^2(1 + \alpha + \beta)}{n\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)(\alpha + \beta + n)} \left[(\alpha + \beta)(\alpha + \beta - 1 + 6n) + 3\alpha\beta(n - 2) + 6n^2 - \frac{3\alpha\beta n(6 - n)}{\alpha + \beta} - \frac{18\alpha\beta n^2}{(\alpha + \beta)^2} \right].$$

Letting $\pi = \frac{\alpha}{\alpha + \beta}$ we note, suggestively, that the mean can be written as

$$\mu = \frac{n\alpha}{\alpha + \beta} = n\pi$$

and the variance as

$$\sigma^2 = \frac{n\alpha\beta(\alpha + \beta + n)}{(\alpha + \beta)^2(\alpha + \beta + 1)} = n\pi(1 - \pi)\frac{\alpha + \beta + n}{\alpha + \beta + 1} = n\pi(1 - \pi)[1 + (n - 1)\rho]$$

where $\rho = \frac{1}{\alpha + \beta + 1}$ is the pairwise correlation between the n Bernoulli draws and is called the over-dispersion parameter.

The following **recurrence relation** holds:

$$\left\{ \begin{array}{l} (\alpha + k)(n - k)p(k) - (k + 1)p(k + 1)(\beta - k + n - 1) = 0, \\ p(0) = \frac{(\beta)_n}{(\alpha + \beta)_n} \end{array} \right\}$$

1.4.3 Point estimates

Method of moments

The **method of moments** estimates can be gained by noting the first and second moments of the beta-binomial namely

$$\mu_1 = \frac{n\alpha}{\alpha + \beta}$$

$$\mu_2 = \frac{n\alpha[n(1 + \alpha) + \beta]}{(\alpha + \beta)(1 + \alpha + \beta)}$$

and setting these raw moments equal to the first and second raw **sample moments** respectively

$$\begin{aligned}\hat{\mu}_1 &= m_1 \\ \hat{\mu}_2 &= m_2\end{aligned}$$

and solving for α and β we get

$$\begin{aligned}\hat{\alpha} &= \frac{nm_1 - m_2}{n(\frac{m_2}{m_1} - m_1 - 1) + m_1} \\ \hat{\beta} &= \frac{(n - m_1)(n - \frac{m_2}{m_1})}{n(\frac{m_2}{m_1} - m_1 - 1) + m_1}.\end{aligned}$$

Note that these estimates can be non-sensically negative which is evidence that the data is either undispersed or under-dispersed relative to the binomial distribution. In this case, the binomial distribution and the **hypergeometric distribution** are alternative candidates respectively.

Maximum likelihood estimation

While closed-form **maximum likelihood estimates** are impractical, given that the pdf consists of common functions (gamma function and/or Beta functions), they can be easily found via direct numerical optimization. Maximum likelihood estimates from empirical data can be computed using general methods for fitting multinomial Pólya distributions, methods for which are described in (Minka 2003). The **R** package VGAM through the function `vglm`, via maximum likelihood, facilitates the fitting of **glm** type models with responses distributed according to the beta-binomial distribution. Note also that there is no requirement that n is fixed throughout the observations.

Example

The following data gives the number of male children among the first 12 children of family size 13 in 6115 families taken from hospital records in 19th century **Saxony** (Sokal and Rohlf, p. 59 from Lindsey). The 13th child is ignored to assuage the effect of families non-randomly stopping when a desired gender is reached.

We note the first two sample moments are

$$\begin{aligned}m_1 &= 6.23 \\ m_2 &= 42.31 \\ n &= 12\end{aligned}$$

and therefore the method of moments estimates are

$$\begin{aligned}\hat{\alpha} &= 34.1350 \\ \hat{\beta} &= 31.6085.\end{aligned}$$

The **maximum likelihood estimates** can be found numerically

$$\begin{aligned}\hat{\alpha}_{\text{mle}} &= 34.09558 \\ \hat{\beta}_{\text{mle}} &= 31.5715\end{aligned}$$

and the maximized log-likelihood is

$$\log \mathcal{L} = -12492.9$$

from which we find the **AIC**

$$AIC = 24989.74.$$

The AIC for the competing binomial model is $AIC = 25070.34$ and thus we see that the beta-binomial model provides a superior fit to the data i.e. there is evidence for overdispersion. **Trivers and Willard** posit a theoretical justification for heterogeneity (also known as "**burstiness**") in gender-proneness among families (i.e. overdispersion).

The superior fit is evident especially among the tails

1.4.4 Further Bayesian considerations

It is convenient to reparameterize the distributions so that the expected mean of the prior is a single parameter: Let

$$\begin{aligned} \pi(\theta|\mu, M) &= \text{Beta}(M\mu, M(1-\mu)) \\ &= \frac{\Gamma(M)}{\Gamma(M\mu)\Gamma(M(1-\mu))} \theta^{M\mu-1} (1-\theta)^{M(1-\mu)-1} \end{aligned}$$

where

$$\begin{aligned} \mu &= \frac{\alpha}{\alpha + \beta} \\ M &= \alpha + \beta \end{aligned}$$

so that

$$\begin{aligned} E(\theta|\mu, M) &= \mu \\ \text{Var}(\theta|\mu, M) &= \frac{\mu(1-\mu)}{M+1}. \end{aligned}$$

The **posterior distribution** $\rho(\theta|k)$ is also a beta distribution:

$$\begin{aligned} \rho(\theta|k) &\propto \ell(k|\theta)\pi(\theta|\mu, M) \\ &= \text{Beta}(k + M\mu, n - k + M(1-\mu)) \\ &= \frac{\Gamma(M)}{\Gamma(M\mu)\Gamma(M(1-\mu))} \binom{n}{k} \theta^{k+M\mu-1} (1-\theta)^{n-k+M(1-\mu)-1} \end{aligned}$$

And

$$E(\theta|k) = \frac{k + M\mu}{n + M}.$$

while the marginal distribution $m(k|\mu, M)$ is given by

$$\begin{aligned} m(k|\mu, M) &= \int_0^1 l(k|\theta)\pi(\theta|\mu, M) d\theta \\ &= \frac{\Gamma(M)}{\Gamma(M\mu)\Gamma(M(1-\mu))} \binom{n}{k} \int_0^1 \theta^{k+M\mu-1} (1-\theta)^{n-k+M(1-\mu)-1} d\theta \\ &= \frac{\Gamma(M)}{\Gamma(M\mu)\Gamma(M(1-\mu))} \binom{n}{k} \frac{\Gamma(k+M\mu)\Gamma(n-k+M(1-\mu))}{\Gamma(n+M)}. \end{aligned}$$

Substituting back M and μ , in terms of α and β , this becomes:

$$m(k|\alpha, \beta) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \frac{\Gamma(k+\alpha)\Gamma(n-k+\beta)}{\Gamma(n+\alpha+\beta)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

which is the expected beta-binomial distribution with parameters n, α and β .

We can also use the method of iterated expectations to find the **expected value** of the marginal moments. Let us write our model as a two-stage compound sampling model. Let k_i be the number of success out of n_i trials for event i :

$$\begin{aligned} k_i &\sim \text{Bin}(n_i, \theta_i) \\ \theta_i &\sim \text{Beta}(\mu, M), \text{ i.i.d.} \end{aligned}$$

We can find iterated moment estimates for the mean and variance using the moments for the distributions in the two-stage model:

$$\begin{aligned} E\left(\frac{k}{n}\right) &= E\left[E\left(\frac{k}{n} \middle| \theta\right)\right] = E(\theta) = \mu \\ \text{var}\left(\frac{k}{n}\right) &= E\left[\text{var}\left(\frac{k}{n} \middle| \theta\right)\right] + \text{var}\left[E\left(\frac{k}{n} \middle| \theta\right)\right] \\ &= E\left[\left(\frac{1}{n}\right)\theta(1-\theta) \middle| \mu, M\right] + \text{var}(\theta|\mu, M) \\ &= \frac{1}{n}(\mu(1-\mu)) + \frac{n-1}{n} \frac{(\mu(1-\mu))}{M+1} \\ &= \frac{\mu(1-\mu)}{n} \left(1 + \frac{n-1}{M+1}\right). \end{aligned}$$

(Here we have used the **law of total expectation** and the **law of total variance**.)

We want point estimates for μ and M . The estimated mean $\hat{\mu}$ is calculated from the sample

$$\hat{\mu} = \frac{\sum_{i=1}^N k_i}{\sum_{i=1}^N n_i}.$$

The estimate of the hyperparameter M is obtained using the moment estimates for the variance of the two-stage model:

$$s^2 = \frac{1}{N} \sum_{i=1}^N \text{var} \left(\frac{k_i}{n_i} \right) = \frac{1}{N} \sum_{i=1}^N \frac{\hat{\mu}(1-\hat{\mu})}{n_i} \left[1 + \frac{n_i - 1}{\widehat{M} + 1} \right]$$

Solving:

$$\widehat{M} = \frac{\hat{\mu}(1-\hat{\mu}) - s^2}{s^2 - \frac{\hat{\mu}(1-\hat{\mu})}{N} \sum_{i=1}^N 1/n_i},$$

where

$$s^2 = \frac{N \sum_{i=1}^N n_i (\hat{\theta}_i - \hat{\mu})^2}{(N-1) \sum_{i=1}^N n_i}.$$

Since we now have parameter point estimates, $\hat{\mu}$ and $\widehat{M}$, for the underlying distribution, we would like to find a point estimate $\tilde{\theta}_i$ for the probability of success for event i . This is the weighted average of the event estimate $\hat{\theta}_i = k_i/n_i$ and $\hat{\mu}$. Given our point estimates for the prior, we may now plug in these values to find a point estimate for the posterior

$$\tilde{\theta}_i = E(\theta|k_i) = \frac{k_i + \widehat{M}\hat{\mu}}{n_i + \widehat{M}} = \frac{\widehat{M}}{n_i + \widehat{M}} \hat{\mu} + \frac{n_i}{n_i + \widehat{M}} \frac{k_i}{n_i}.$$

1.4.5 Shrinkage factors

We may write the posterior estimate as a weighted average:

$$\tilde{\theta}_i = \hat{B}_i \hat{\mu} + (1 - \hat{B}_i) \hat{\theta}_i$$

where $\hat{B}_i$ is called the *shrinkage factor*.

$$\hat{B}_i = \frac{\widehat{M}}{\widehat{M} + n_i}$$

1.4.6 Related distributions

- $BB(1, 1, n) \sim U(0, n)$ where $U(a, b)$ is the **discrete uniform distribution**.

1.4.7 See also

- **Dirichlet-multinomial distribution**

1.4.8 References

- Minka, Thomas P. (2003). *Estimating a Dirichlet distribution*. Microsoft Technical Report.

1.4.9 External links

- Using the Beta-binomial distribution to assess performance of a biometric identification device
- Fastfit contains Matlab code for fitting Beta-Binomial distributions (in the form of two-dimensional Pólya distributions) to data.
- Interactive graphic: [Univariate Distribution Relationships](#)
- Beta-Binomial distribution package for R

1.5 Degenerate distribution

In mathematics, a **degenerate distribution** or **deterministic distribution** is the probability distribution of a random variable which only takes a single value. Examples include a two-headed coin and rolling a die whose sides all show the same number. This distribution satisfies the definition of “random variable” even though it does not appear random in the everyday sense of the word; hence it is considered *degenerate*.

The degenerate distribution is localized at a point k_0 on the real line. The probability mass function equals 1 at this point and 0 elsewhere.

The distribution can be viewed as the limiting case of a continuous distribution whose variance goes to 0 causing the probability density function to be a *delta function* at k_0 , with infinite height there but area equal to 1.

The cumulative distribution function of the degenerate distribution is:

$$F(k; k_0) = \begin{cases} 1, & \text{if } k \geq k_0 \\ 0, & \text{if } k < k_0 \end{cases}$$

1.5.1 Constant random variable

In probability theory, a **constant random variable** is a discrete random variable that takes a constant value, regardless of any event that occurs. This is technically different from an **almost surely constant random variable**, which may take other values, but only on events with probability zero. Constant and almost surely constant random variables provide a way to deal with constant values in a probabilistic framework.

Let $X: \Omega \rightarrow \mathbf{R}$ be a random variable defined on a probability space (Ω, P) . Then X is an *almost surely constant random variable* if there exists $c \in \mathbb{R}$ such that

$$\Pr(X = c) = 1,$$

and is furthermore a *constant random variable* if

$$X(\omega) = c, \quad \forall \omega \in \Omega.$$

Note that a constant random variable is almost surely constant, but not necessarily *vice versa*, since if X is almost surely constant then there may exist $\gamma \in \Omega$ such that $X(\gamma) \neq c$ (but then necessarily $\Pr(\{\gamma\}) = 0$, in fact $\Pr(X \neq c) = 0$).

For practical purposes, the distinction between X being constant or almost surely constant is unimportant, since the cumulative distribution function $F(x)$ of X does not depend on whether X is constant or 'merely' almost surely constant. In this case,

$$F(x) = \begin{cases} 1, & x \geq c, \\ 0, & x < c. \end{cases}$$

The function $F(x)$ is a **step function**; in particular it is a translation of the **Heaviside step function**.

1.6 Hypergeometric distribution

In **probability theory** and **statistics**, the **hypergeometric distribution** is a discrete probability distribution that describes the probability of k successes in n draws, *without* replacement, from a finite **population** of size N that contains exactly K successes, wherein each draw is either a success or a failure. In contrast, the **binomial distribution** describes the probability of k successes in n draws *with* replacement.

In **statistics**, the **hypergeometric test** uses the hypergeometric distribution to calculate the statistical significance of having drawn a specific k successes (out of n total draws) from the aforementioned population. The test is often used to identify which sub-populations are over- or under-represented in a sample. This test has a wide range of applications. For example, a marketing group could use the test to understand their customer base by testing a set of known customers for over-representation of various demographic subgroups (e.g., women, people under 30).

1.6.1 Definition

The following conditions characterize the hypergeometric distribution:

- The result of each draw (the elements of the population being sampled) can be classified into one of two mutually exclusive categories (e.g. Pass/Fail or Female/Male or Employed/Unemployed).
- The probability of a success changes on each draw, as each draw decreases the population (*sampling without replacement* from a finite population).

A **random variable** X follows the hypergeometric distribution if its **probability mass function** (pmf) is given by^[1]

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

where

- N is the population size,
- K is the number of success states in the population,
- n is the number of draws,
- k is the number of observed successes,
- $\binom{a}{b}$ is a **binomial coefficient**.

The pmf is positive when $\max(0, n + K - N) \leq k \leq \min(K, n)$.

The pmf satisfies the **recurrence relation**

$$(k+1)(N-K-(n-k-1))P(X=k+1) = (K-k)(n-k)P(X=k)$$

with

$$P(X = 0) = \frac{\binom{N-K}{n}}{\binom{N}{n}}$$

1.6.2 Combinatorial identities

As one would expect, the probabilities sum up to 1 :

$$\sum_{0 \leq k \leq n} \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = 1$$

This is essentially **Vandermonde's identity** from combinatorics.

Also note the following identity holds:

$$\frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\binom{n}{k} \binom{N-n}{K-k}}{\binom{N}{K}}.$$

This follows from the symmetry of the problem, but it can also be shown by expressing the binomial coefficients in terms of factorials and rearranging the latter.

1.6.3 Application and example

The classical application of the hypergeometric distribution is **sampling without replacement**. Think of an **urn** with two types of **marbles**, red ones and green ones. Define drawing a green marble as a success and drawing a red marble as a failure (analogous to the binomial distribution). If the variable N describes the number of **all marbles in the urn** (see contingency table below) and K describes the number of **green marbles**, then $N - K$ corresponds to the number of **red marbles**. In this example, X is the **random variable** whose outcome is k , the number of green marbles actually drawn in the experiment. This situation is illustrated by the following **contingency table**:

Now, assume (for example) that there are 5 green and 45 red marbles in the urn. Standing next to the urn, you close your eyes and draw 10 marbles without replacement. What is the probability that exactly 4 of the 10 are green? *Note that although we are looking at success/failure, the data are not accurately modeled by the **binomial distribution**, because the probability of success on each trial is not the same, as the size of the remaining population changes as we remove each marble.*

This problem is summarized by the following contingency table:

The probability of drawing exactly k green marbles can be calculated by the formula

$$P(X = k) = f(k; N, K, n) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}.$$

Hence, in this example calculate

$$P(X = 4) = f(4; 50, 5, 10) = \frac{\binom{5}{4} \binom{45}{6}}{\binom{50}{10}} = \frac{5 \cdot 8145060}{10272278170} = 0.003964583 \dots$$

Intuitively we would expect it to be even more unlikely for all 5 marbles to be green.

$$P(X = 5) = f(5; 50, 5, 10) = \frac{\binom{5}{5} \binom{45}{5}}{\binom{50}{10}} = \frac{1 \cdot 1221759}{10272278170} = 0.0001189375 \dots,$$

As expected, the probability of drawing 5 green marbles is roughly 35 times less likely than that of drawing 4.

Application to Texas Hold'em Poker

In **Hold'em** Poker players make the best hand they can combining the two cards in their hand with the 5 cards (community cards) eventually turned up on the table. The deck has 52 and there are 13 of each suit. For this example assume a player has 2 clubs in the hand and there are 3 cards showing on the table, 2 of which are also clubs. The player would like to know the probability of one of the next 2 cards to be shown being a club to complete his **flush**.

There are 4 clubs showing so there are 9 still unseen. There are 5 cards showing (2 in the hand and 3 on the table) so there are $52 - 5 = 47$ still unseen.

The probability that one of the next two cards turned is a club can be calculated using hypergeometric with $k = 1, n = 2, K = 9$ and $N = 47$. (about 31.6%)

The probability that both of the next two cards turned are clubs can be calculated using hypergeometric with $k = 2, n = 2, K = 9$ and $N = 47$. (about 3.3%)

The probability that neither of the next two cards turned are clubs can be calculated using hypergeometric with $k = 0, n = 2, K = 9$ and $N = 47$. (about 65.0%)

1.6.4 Symmetries

Swapping the roles of green and red marbles:

$$f(k; N, K, n) = f(n - k; N, N - K, n)$$

Swapping the roles of drawn and not drawn marbles:

$$f(k; N, K, n) = f(K - k; N, K, N - n)$$

Swapping the roles of green and drawn marbles:

$$f(k; N, K, n) = f(k; N, n, K)$$

1.6.5 Hypergeometric test

The **hypergeometric test** uses the hypergeometric distribution to measure the statistical significance of having drawn a sample consisting of a specific number of k successes (out of n total draws) from a population of size N containing K successes. In a test for over-representation of successes in the sample, the hypergeometric p-value is calculated as the probability of randomly drawing k or more successes from the population in n total draws. In a test for under-representation, the p-value is the probability of randomly drawing k or fewer successes.

Relationship to Fisher's exact test

See also: **Fisher's noncentral hypergeometric distribution**

The test based on the hypergeometric distribution (hypergeometric test) is identical to the corresponding one-tailed version of **Fisher's exact test**^[2]. Reciprocally, the p-value of a two-sided Fisher's exact test can be calculated as the sum of two appropriate hypergeometric tests (for more information see^[3]).

1.6.6 Order of draws

The probability of drawing any sequence of white and black marbles (the hypergeometric distribution) depends only on the number of white and black marbles, not on the order in which they appear; i.e., it is an **exchangeable** distribution. As a result, the probability of drawing a white marble in the i^{th} draw is^[4]

$$P(W_i) = \frac{K}{N}.$$

1.6.7 Related distributions

Let $X \sim \text{Hypergeometric}(K, N, n)$ and $p = K/N$.

- If $n = 1$ then X has a **Bernoulli distribution** with parameter p .
- Let Y have a **binomial distribution** with parameters n and p ; this models the number of successes in the analogous sampling problem *with* replacement. If N and K are large compared to n , and p is not close to 0 or 1, then X and Y have similar distributions, i.e., $P(X \leq k) \approx P(Y \leq k)$.
- If n is large, N and K are large compared to n , and p is not close to 0 or 1, then

$$P(X \leq k) \approx \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right)$$

where Φ is the **standard normal distribution function**

- If the probabilities to draw a white or black marble are not equal (e.g. because white marbles are bigger/easier to grasp than black marbles) then X has a **noncentral hypergeometric distribution**
- The **beta-binomial distribution** is a **conjugate prior** for the hypergeometric distribution.

The following table describes four distributions related to the number of successes in a sequence of draws:

1.6.8 Multivariate hypergeometric distribution

The model of an **urn** with black and white marbles can be extended to the case where there are more than two colors of marbles. If there are K_i marbles of color i in the urn and you take n marbles at random without replacement, then the number of marbles of each color in the sample $(k_1, k_2, \dots, k_c)$ has the multivariate hypergeometric distribution. This has the same relationship to the **multinomial distribution** that the hypergeometric distribution has to the binomial distribution—the multinomial distribution is the “with-replacement” distribution and the multivariate hypergeometric is the “without-replacement” distribution.

The properties of this distribution are given in the adjacent table, where c is the number of different colors and $N = \sum_{i=1}^c K_i$ is the total number of marbles.

Example

Suppose there are 5 black, 10 white, and 15 red marbles in an urn. You reach in and randomly select six marbles without replacement. What is the probability that you pick exactly two of each color?

$$P(2\text{black}, 2\text{white}, 2\text{red}) = \frac{\binom{5}{2} \binom{10}{2} \binom{15}{2}}{\binom{30}{6}} = 0.079575596816976$$

Note: When picking the six marbles without replacement, the expected number of black marbles is $6 \times (5/30) = 1$, the expected number of white marbles is $6 \times (10/30) = 2$, and the expected number of red marbles is $6 \times (15/30) = 3$. This comes from the expected value of a Binomial distribution, $E(X) = np$.

1.6.9 See also

- Multinomial distribution
- Sampling (statistics)
- Generalized hypergeometric function
- Coupon collector's problem
- Geometric distribution
- Keno

1.6.10 Notes

- [1] Rice, John A. (2007). *Mathematical Statistics and Data Analysis* (Third ed.). Duxbury Press. p. 42.
- [2] Rivals, I.; Personnaz, L.; Taing, L.; Potier, M.-C (2007). "Enrichment or depletion of a GO category within a class of genes: which test?". *Bioinformatics* **23** (4): 401–407. doi:10.1093/bioinformatics/btl633. PMID 17182697.
- [3] K. Preacher and N. Briggs. "Calculation for Fisher's Exact Test: An interactive calculation tool for Fisher's exact probability test for 2 x 2 tables (interactive page)".
- [4] <http://www.stat.yale.edu/~jpollard/Courses/600.spring2010/Handouts/Symmetry%5BPolyaUrn%5D.pdf>

1.6.11 References

- Berkopeć, Aleš (2007). "HyperQuick algorithm for discrete hypergeometric distribution". *Journal of Discrete Algorithms* **5** (2): 341. doi:10.1016/j.jda.2006.01.001.
- Skala, M. (2011). "Hypergeometric tail inequalities: ending the insanity" (PDF). unpublished note

1.6.12 External links

- The Hypergeometric Distribution and Binomial Approximation to a Hypergeometric Random Variable by Chris Boucher, Wolfram Demonstrations Project.
- Weisstein, Eric W., "Hypergeometric Distribution", *Math World*.

1.7 Poisson binomial distribution

In probability theory and statistics, the **Poisson binomial distribution** is the discrete probability distribution of a sum of independent Bernoulli trials that are not necessarily identically distributed. The concept is named after Siméon Denis Poisson.

In other words, it is the probability distribution of the number of successes in a sequence of n independent yes/no experiments with success probabilities $p_1, p_2, \dots, p_n$. The ordinary binomial distribution is a special case of the Poisson binomial distribution, when all success probabilities are the same, that is $p_1 = p_2 = \dots = p_n$.

1.7.1 Mean and variance

Since a Poisson binomial distributed variable is a sum of n independent Bernoulli distributed variables, its mean and variance will simply be sums of the mean and variance of the n Bernoulli distributions:

$$\mu = \sum_{i=1}^n p_i$$

$$\sigma^2 = \sum_{i=1}^n (1 - p_i)p_i$$

For fixed values of the mean (μ) and size (n), the variance is maximal when all success probabilities are equal and we have a binomial distribution. When the mean is fixed, the variance is bounded from above by the variance of the Poisson distribution with the same mean which is attained asymptotically as n tends to infinity.

1.7.2 Probability mass function

The probability of having k successful trials out of a total of n can be written as the sum ^[1]

$$\Pr(K = k) = \sum_{A \in F_k} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j)$$

where F_k is the set of all subsets of k integers that can be selected from $\{1, 2, 3, \dots, n\}$. For example, if $n = 3$, then $F_2 = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. A^c is the complement of A , i.e. $A^c = \{1, 2, 3, \dots, n\} \setminus A$.

F_k will contain $n! / ((n - k)!k!)$ elements, the sum over which is infeasible to compute in practice unless the number of trials n is small (e.g. if $n = 30$, F_{15} contains over 10^{20} elements). However, there are other, more efficient ways to calculate $\Pr(K = k)$.

As long as none of the success probabilities are equal to one, one can calculate the probability of k successes using the recursive formula ^{[2] [3]}

$$\Pr(K = k) = \begin{cases} \prod_{i=1}^n (1 - p_i) & k = 0 \\ \frac{1}{k} \sum_{i=1}^k (-1)^{i-1} \Pr(K = k - i) T(i) & k > 0 \end{cases}$$

where

$$T(i) = \sum_{j=1}^n \left(\frac{p_j}{1 - p_j} \right)^i.$$

The recursive formula is not numerically stable, and should be avoided if n is greater than approximately 20. Another possibility is using the **discrete Fourier transform** ^[4]

$$\Pr(K = k) = \frac{1}{n+1} \sum_{l=0}^n C^{-lk} \prod_{m=1}^n (1 + (C^l - 1)p_m)$$

where $C = \exp\left(\frac{2i\pi}{n+1}\right)$ and $i = \sqrt{-1}$.

Still other methods are described in ^[5]

1.7.3 Entropy

There is no simple formula for the entropy of a Poisson binomial distribution, but the entropy can be upper bounded by that entropy of a binomial distribution with the same number parameter and the same mean. Therefore the entropy can also be upper bounded by the entropy of a Poisson distribution with the same mean. ^[6]

The Shepp-Olkin conjecture, due to **Lawrence Shepp** and **Ingram Olkin** in 1981, states that the entropy of a Poisson binomial distribution is a concave function of the success probabilities $p_1, p_2, \dots, p_n$. ^[7]

1.7.4 See also

- **Le Cam's theorem**
- **Binomial distribution**
- **Poisson distribution**

1.7.5 References

- [1] Wang, Y. H. (1993). "On the number of successes in independent trials" (PDF). *Statistica Sinica* **3** (2): 295–312.
- [2] Shah, B. K. (1994). "On the distribution of the sum of independent integer valued random variables". *American Statistician* **27** (3): 123–124. JSTOR 2683639.
- [3] Chen, X. H.; A. P. Dempster; J. S. Liu (1994). "Weighted finite population sampling to maximize entropy" (PDF). *Biometrika* **81** (3): 457. doi:10.1093/biomet/81.3.457.
- [4] Fernandez, M.; S. Williams (2010). "Closed-Form Expression for the Poisson-Binomial Probability Density Function". *IEEE Transactions on Aerospace Electronic Systems* **46**: 803–817. doi:10.1109/TAES.2010.5461658.
- [5] Chen, S. X.; J. S. Liu (1997). "Statistical Applications of the Poisson-Binomial and conditional Bernoulli distributions". *Statistica Sinica* **7**: 875–892.
- [6] Harremoës, P. (2001). "Binomial and Poisson distributions as maximum entropy distributions" (PDF). *IEEE Transactions on Information Theory* **47** (5): 2039–2041. doi:10.1109/18.930936.
- [7] Shepp, Lawrence; Olkin, Ingram (1981). "Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution". In Gani, J.; Rohatgi, V.K. *Contributions to probability: A collection of papers dedicated to Eugene Lukacs*. New York,: Academic Press,. pp. 201–206. ISBN 0-12-274460-8. MR 0618689.

1.8 Fisher's noncentral hypergeometric distribution

In probability theory and statistics, **Fisher's noncentral hypergeometric distribution** is a generalization of the **hypergeometric distribution** where sampling probabilities are modified by weight factors. It can also be defined as the **conditional distribution** of two or more **binomially distributed** variables dependent upon their fixed sum.

The distribution may be illustrated by the following **urn model**. Assume, for example, that an urn contains m_1 red balls and m_2 white balls, totalling $N = m_1 + m_2$ balls. Each red ball has the weight ω_1 and each white ball has the weight ω_2 . We will say that the odds ratio is $\omega = \omega_1 / \omega_2$. Now we are taking balls randomly in such a way that the probability of taking a particular ball is proportional to its weight, but independent of what happens to the other balls. The number of balls taken of a particular color follows the **binomial distribution**. If the total number n of balls taken is known then the conditional distribution of the number of taken red balls for given n is Fisher's noncentral hypergeometric distribution. To generate this distribution experimentally, we have to repeat the experiment until it happens to give n balls.

If we want to fix the value of n prior to the experiment then we have to take the balls one by one until we have n balls. The balls are therefore no longer independent. This gives a slightly different distribution known as **Wallenius' noncentral hypergeometric distribution**. It is far from obvious why these two distributions are different. See the entry for **noncentral hypergeometric distributions** for an explanation of the difference between these two distributions and a discussion of which distribution to use in various situations.

The two distributions are both equal to the (central) **hypergeometric distribution** when the odds ratio is 1.

Unfortunately, both distributions are known in the literature as “the” noncentral hypergeometric distribution. It is important to be specific about which distribution is meant when using this name.

Fisher's noncentral hypergeometric distribution was first given the name **extended hypergeometric distribution** (Harkness, 1965), and some authors still use this name today.

1.8.1 Univariate distribution

The probability function, mean and variance are given in the table to the right.

An alternative expression of the distribution has both the number of balls taken of each color and the number of balls not taken as random variables, whereby the expression for the probability becomes symmetric.

The calculation time for the probability function can be high when the sum in P_0 has many terms. The calculation time can be reduced by calculating the terms in the sum recursively relative to the term for $y = x$ and ignoring negligible terms in the tails (Liao and Rosen, 2001).

The mean can be approximated by:

$$\mu \approx \frac{-2c}{b - \sqrt{b^2 - 4ac}}$$

where $a = \omega - 1$, $b = m_1 + n - N - (m_1 + n)\omega$, $c = m_1 n \omega$.

The variance can be approximated by:

$$\sigma^2 \approx \frac{N}{N-1} \left/ \left(\frac{1}{\mu} + \frac{1}{m_1 - \mu} + \frac{1}{n - \mu} + \frac{1}{\mu + m_2 - n} \right) \right.$$

Better approximations to the mean and variance are given by Levin (1984, 1990), McCullagh and Nelder (1989), Liao (1992), and Eisinga and Pelzer (2011). The saddlepoint methods to approximate the mean and the variance suggested Eisinga and Pelzer (2011) offer extremely accurate results.

Properties

The following symmetry relations apply:

$$\text{fnchypg}(x; n, m_1, N, \omega) = \text{fnchypg}(n - x; n, m_2, N, 1/\omega).$$

$$\text{fnchypg}(x; n, m_1, N, \omega) = \text{fnchypg}(x; m_1, n, N, \omega).$$

$$\text{fnchypg}(x; n, m_1, N, \omega) = \text{fnchypg}(m_1 - x; N - n, m_1, N, 1/\omega).$$

Recurrence relation:

$$\text{fnchypg}(x; n, m_1, N, \omega) = \text{fnchypg}(x - 1; n, m_1, N, \omega) \frac{(m_1 - x + 1)(n - x + 1)}{x(m_2 - n + x)} \omega.$$

Recurrence relation

A Fisher hypergeometric distribution gives the distribution of the number of successes in n independent draws from a population of size n_{tot} containing n_{succ} successes with the odds ratio w .

$$\left\{ \begin{array}{l} wf(x)(x - n)(n_{\text{succ}} - x) - (x + 1)f(x + 1)(n + n_{\text{succ}} - n_{\text{tot}} - x - 1) = 0, \\ f(0) = \frac{1}{{}_2F_1(-n, -n_{\text{succ}}; -n - n_{\text{succ}} + n_{\text{tot}} + 1; w)} \end{array} \right\}$$

1.8.2 Multivariate distribution

The distribution can be expanded to any number of colors c of balls in the urn. The multivariate distribution is used when there are more than two colors.

The probability function and a simple approximation to the mean are given to the right. Better approximations to the mean and variance are given by McCullagh and Nelder (1989).

Properties

The order of the colors is arbitrary so that any colors can be swapped.

The weights can be arbitrarily scaled:

$$\text{mfncchypg}(\mathbf{x}; n, \mathbf{m}, \boldsymbol{\omega}) = \text{mfncchypg}(\mathbf{x}; n, \mathbf{m}, r\boldsymbol{\omega}) \text{ for all } r \in \mathbb{R}_+.$$

Colors with zero number ($m_i = 0$) or zero weight ($\omega_i = 0$) can be omitted from the equations.

Colors with the same weight can be joined:

$$\begin{aligned} & \text{mfncchypg}(\mathbf{x}; n, \mathbf{m}, (\omega_1, \dots, \omega_{c-1}, \omega_{c-1})) \\ &= \text{mfncchypg}((x_1, \dots, x_{c-1} + x_c); n, (m_1, \dots, m_{c-1} + m_c), (\omega_1, \dots, \omega_{c-1})) \cdot \\ & \quad \text{hypg}(x_c; x_{c-1} + x_c, m_c, m_{c-1} + m_c) \end{aligned}$$

where $\text{hypg}(x; n, m, N)$ is the (univariate, central) hypergeometric distribution probability.

1.8.3 Applications

Fisher's noncentral hypergeometric distribution is useful for models of biased sampling or biased selection where the individual items are sampled independently of each other with no competition. The bias or odds can be estimated from an experimental value of the mean. Use Wallenius' noncentral hypergeometric distribution instead if items are sampled one by one with competition.

Fisher's noncentral hypergeometric distribution is used mostly for tests in contingency tables where a conditional distribution for fixed margins is desired. This can be useful, for example, for testing or measuring the effect of a medicine. See McCullagh and Nelder (1989).

1.8.4 Software available

- FisherHypergeometricDistribution in Mathematica.
- An implementation for the R programming language is available as the package named BiasedUrn. Includes univariate and multivariate probability mass functions, distribution functions, quantiles, random variable generating functions, mean and variance.
- The R package MCMCpack includes the univariate probability mass function and random variable generating function.
- SAS System includes univariate probability mass function and distribution function.
- Implementation in C++ is available from www.agner.org.
- Calculation methods are described by Liao and Rosen (2001) and Fog (2008).

1.8.5 See also

- Noncentral hypergeometric distributions
- Wallenius' noncentral hypergeometric distribution
- Hypergeometric distribution
- Urn models
- Biased sample
- Bias
- Contingency table
- Fisher's exact test

1.8.6 References

Breslow, N. E.; Day, N. E. (1980), *Statistical Methods in Cancer Research*, Lyon: International Agency for Research on Cancer.

Eisinga, R.; Pelzer, B. (2011), "Saddlepoint approximations to the mean and variance of the extended hypergeometric distribution", *Statistica Neerlandica* **65** (1): 22–31, doi:10.1111/j.1467-9574.2010.00468.x.

Fog, A. (2007), *Random number theory*.

Fog, A. (2008), "Sampling Methods for Wallenius' and Fisher's Noncentral Hypergeometric Distributions", *Communications in statistics, Simulation and Computation* **37** (2): 241–257, doi:10.1080/03610910701790236.

- Johnson, N. L.; Kemp, A. W.; Kotz, S. (2005), *Univariate Discrete Distributions*, Hoboken, New Jersey: Wiley and Sons.
- Levin, B. (1984), "Simple Improvements on Cornfield's approximation to the mean of a noncentral Hypergeometric random variable", *Biometrika* **71** (3): 630–632, doi:10.1093/biomet/71.3.630.
- Levin, B. (1990), "The saddlepoint correction in conditional logistic likelihood analysis", *Biometrika* **77** (2): 275–285, JSTOR 2336805.
- Liao, J. (1992), "An Algorithm for the Mean and Variance of the Noncentral Hypergeometric Distribution", *Biometrics* **48** (3): 889–892, doi:10.2307/2532354, JSTOR 2532354.
- Liao, J. G.; Rosen, O. (2001), "Fast and Stable Algorithms for Computing and Sampling from the Noncentral Hypergeometric Distribution", *The American Statistician* **55** (4): 366–369, doi:10.1198/000313001753272547.
- McCullagh, P.; Nelder, J. A. (1989), *Generalized Linear Models*, 2. ed., London: Chapman and Hall.

1.9 Wallenius' noncentral hypergeometric distribution

In probability theory and statistics, **Wallenius' noncentral hypergeometric distribution** (named after Kenneth Ted Wallenius) is a generalization of the **hypergeometric distribution** where items are sampled with bias.

This distribution can be illustrated as an **urn model** with bias. Assume, for example, that an urn contains m_1 red balls and m_2 white balls, totalling $N = m_1 + m_2$ balls. Each red ball has the weight ω_1 and each white ball has the weight ω_2 . We will say that the odds ratio is $\omega = \omega_1 / \omega_2$. Now we are taking n balls, one by one, in such a way that the probability of taking a particular ball at a particular draw is equal to its proportion of the total weight of all balls that lie in the urn at that moment. The number of red balls x_1 that we get in this experiment is a random variable with Wallenius' noncentral hypergeometric distribution.

The matter is complicated by the fact that there is more than one noncentral hypergeometric distribution. Wallenius' noncentral hypergeometric distribution is obtained if balls are sampled one by one in such a way that there is competition between the balls. Fisher's noncentral hypergeometric distribution is obtained if the balls are sampled simultaneously or independently of each other. Unfortunately, both distributions are known in the literature as "the" noncentral hypergeometric distribution. It is important to be specific about which distribution is meant when using this name.

The two distributions are both equal to the (central) **hypergeometric distribution** when the odds ratio is 1.

It is far from obvious why these two distributions are different. See the Wikipedia entry on **noncentral hypergeometric distributions** for a more detailed explanation of the difference between these two probability distributions.

1.9.1 Univariate distribution

Wallenius' distribution is particularly complicated because each ball has a probability of being taken that depends not only on its weight, but also on the total weight of its competitors. And the weight of the competing balls depends on the outcomes of all preceding draws.

This recursive dependency gives rise to a **difference equation** with a solution that is given in **open form** by the integral in the expression of the probability mass function in the table above.

Closed form expressions for the probability mass function exist (Lyons, 1980), but they are not very useful for practical calculations because of extreme **numerical instability**, except in degenerate cases.

Several other calculation methods are used, including **recursion**, **Taylor expansion** and **numerical integration** (Fog, 2007, 2008).

The most reliable calculation method is recursive calculation of $f(x, n)$ from $f(x, n-1)$ and $f(x-1, n-1)$ using the recursion formula given below under properties. The probabilities of all (x, n) combinations on all possible **trajectories** leading to the desired point are calculated, starting with $f(0, 0) = 1$ as shown on the figure to the right. The total number of probabilities to calculate is $n(x+1) - x^2$. Other calculation methods must be used when n and x are so big that this method is too inefficient.

The probability that all balls have the same color is easier to calculate. See the formula below under multivariate distribution.

No exact formula for the mean is known (short of complete enumeration of all probabilities). The equation given above is reasonably accurate. This equation can be solved for μ by **Newton-Raphson iteration**. The same equation can be used for estimating the odds from an experimentally obtained value of the mean.

Properties of the univariate distribution

Wallenius' distribution has fewer symmetry relations than **Fisher's noncentral hypergeometric distribution** has. The only symmetry relates to the swapping of colors:

$$\text{wnchypg}(x; n, m_1, m_2, \omega) = \text{wnchypg}(n - x; n, m_2, m_1, 1/\omega) .$$

Unlike Fisher's distribution, Wallenius' distribution has no symmetry relating to the number of balls *not* taken.

The following recursion formula is useful for calculating probabilities:

$$\begin{aligned} \text{wnchypg}(x; n, m_1, m_2, \omega) = \\ \text{wnchypg}(x - 1; n - 1, m_1, m_2, \omega) \frac{(m_1 - x + 1)\omega}{(m_1 - x + 1)\omega + m_2 + x - n} + \\ \text{wnchypg}(x; n - 1, m_1, m_2, \omega) \frac{m_2 + x - n + 1}{(m_1 - x)\omega + m_2 + x - n + 1} \end{aligned}$$

Another recursion formula is also known:

$$\begin{aligned} \text{wnchypg}(x; n, m_1, m_2, \omega) = \\ \text{wnchypg}(x - 1; n - 1, m_1 - 1, m_2, \omega) \frac{m_1\omega}{m_1\omega + m_2} + \\ \text{wnchypg}(x; n - 1, m_1, m_2 - 1, \omega) \frac{m_2}{m_1\omega + m_2} . \end{aligned}$$

The probability is limited by

$$f_1(x) \leq \text{wnchypg}(x; n, m_1, m_2, \omega) \leq f_2(x) , \text{ for } \omega < 1 ,$$

$$f_1(x) \geq \text{wnchypg}(x; n, m_1, m_2, \omega) \geq f_2(x) , \text{ for } \omega > 1 , \text{ where}$$

$$f_1(x) = \binom{m_1}{x} \binom{m_2}{n-x} \frac{n!}{(m_1 + m_2/\omega)^{\underline{x}} (m_2 + \omega(m_1 - x))^{\underline{n-x}}}$$

$$f_2(x) = \binom{m_1}{x} \binom{m_2}{n-x} \frac{n!}{(m_1 + (m_2 - x_2)/\omega)^{\underline{x}} (m_2 + \omega m_1)^{\underline{n-x}}} ,$$

where the underlined superscript indicates the **falling factorial** $a^{\underline{b}} = a(a-1) \dots (a-b+1) .$

1.9.2 Multivariate distribution

The distribution can be expanded to any number of colors c of balls in the urn. The multivariate distribution is used when there are more than two colors.

The probability mass function can be calculated by various **Taylor expansion** methods or by **numerical integration** (Fog, 2008).

The probability that all balls have the same color, j , can be calculated as:

$$\text{mwnchypg}((0, \dots, 0, x_j, 0, \dots); n, \mathbf{m}, \boldsymbol{\omega}) = \frac{m_j^{\underline{n}}}{\left(\frac{1}{\omega_j} \sum_{i=1}^c m_i \omega_i\right)^{\underline{n}}}$$

for $x_j = n \leq m_j$, where the underlined superscript denotes the **falling factorial**.

A reasonably good approximation to the mean can be calculated using the equation given above. The equation can be solved by defining θ so that

$$\mu_i = m_i(1 - e^{\omega_i \theta})$$

and solving

$$\sum_{i=1}^c \mu_i = n$$

for θ by **Newton-Raphson** iteration.

The equation for the mean is also useful for estimating the odds from experimentally obtained values for the mean.

No good way of calculating the variance is known. The best known method is to approximate the multivariate Wallenius distribution by a multivariate **Fisher's noncentral hypergeometric distribution** with the same mean, and insert the mean as calculated above in the approximate formula for the variance of the latter distribution.

Properties of the multivariate distribution

The order of the colors is arbitrary so that any colors can be swapped.

The weights can be arbitrarily scaled:

$$\text{mwnchypg}(\mathbf{x}; n, \mathbf{m}, \boldsymbol{\omega}) = \text{mwnchypg}(\mathbf{x}; n, \mathbf{m}, r\boldsymbol{\omega}) \quad \text{for all } r \in \mathbb{R}_+.$$

Colors with zero number ($m_i = 0$) or zero weight ($\omega_i = 0$) can be omitted from the equations.

Colors with the same weight can be joined:

$$\begin{aligned} \text{mwnchypg}(\mathbf{x}; n, \mathbf{m}, (\omega_1, \dots, \omega_{c-1}, \omega_{c-1})) &= \\ \text{mwnchypg}((x_1, \dots, x_{c-1} + x_c); n, (m_1, \dots, m_{c-1} + m_c), (\omega_1, \dots, \omega_{c-1})) &\cdot \\ \text{hypg}(x_c; x_{c-1} + x_c, m_c, m_{c-1} + m_c), & \end{aligned}$$

where $\text{hypg}(x; n, m, N)$ is the (univariate, central) hypergeometric distribution probability.

1.9.3 Complementary Wallenius' noncentral hypergeometric distribution

The balls that are *not* taken in the urn experiment have a distribution that is different from Wallenius' noncentral hypergeometric distribution, due to a lack of symmetry. The distribution of the balls not taken can be called the **complementary Wallenius' noncentral hypergeometric distribution**.

Probabilities in the complementary distribution are calculated from Wallenius' distribution by replacing n with $N-n$, x_i with $m_i - x_i$, and ω_i with $1/\omega_i$.

1.9.4 Software available

- WalleniusHypergeometricDistribution in Mathematica.
- An implementation for the R programming language is available as the package named BiasedUrn. Includes univariate and multivariate probability mass functions, distribution functions, quantiles, random variable generating functions, mean and variance.
- Implementation in C++ is available from www.agner.org.

1.9.5 See also

- Noncentral hypergeometric distributions
- Fisher's noncentral hypergeometric distribution
- Biased sample
- Bias
- Population genetics
- Fisher's exact test

1.9.6 References

- Chesson, J. (1976), "A non-central multivariate hypergeometric distribution arising from biased sampling with application to selective predation", *Journal of Applied Probability* (Applied Probability Trust) **13** (4): 795–797, doi:10.2307/3212535, JSTOR 3212535.
- Fog, A. (2007), *Random number theory*.
- Fog, A. (2008), "Calculation Methods for Wallenius' Noncentral Hypergeometric Distribution", *Communications in statistics, Simulation and Computation* **37** (2): 258–273, doi:10.1080/03610910701790269.
- Johnson, N. L.; Kemp, A. W.; Kotz, S. (2005), *Univariate Discrete Distributions*, Hoboken, New Jersey: Wiley and Sons.
- Lyons, N. I. (1980), "Closed Expressions for Noncentral Hypergeometric Probabilities", *Communications in statistics, B* **9**: 313–314.
- Manly, B. F. J. (1974), "A Model for Certain Types of Selection Experiments", *Biometrics* (International Biometric Society) **30** (2): 281–294, doi:10.2307/2529649, JSTOR 2529649.
- Wallenius, K. T. (1963), *Biased Sampling: The Non-central Hypergeometric Probability Distribution*. Ph.D. Thesis, Stanford University, Department of Statistics.

1.10 Benford's law

Not to be confused with the unrelated adage **Benford's law of controversy**.

Benford's law, also called the **First-Digit Law**, is a **phenomenological** law about the frequency distribution of leading digits in many (but not all) real-life sets of numerical data. That law states that in many naturally occurring collections of numbers the small digits occur disproportionately often as leading **significant digits**.^[1] For example, in sets which obey the law the number 1 would appear as the most significant digit about 30% of the time, while larger digits would occur in that position less frequently: 9 would appear less than 5% of the time. If all digits were distributed uniformly, they would each occur about 11.1% of the time.^[2] Benford's law also concerns the expected distribution for digits beyond the first, which approach a uniform distribution.

It has been shown that this result applies to a wide variety of data sets, including electricity bills, street addresses, stock prices, population numbers, death rates, lengths of rivers, **physical** and **mathematical constants**,^[3] and processes described by **power laws** (which are very common in nature). It tends to be most accurate when values are distributed across multiple orders of magnitude.

The graph here shows Benford's law for **base 10**. There is a generalization of the law to numbers expressed in other bases (for example, **base 16**), and also a generalization from leading 1 digit to leading n digits.

It is named after physicist **Frank Benford**, who stated it in 1938,^[4] although it had been previously stated by **Simon Newcomb** in 1881.^[5]

1.10.1 Mathematical statement

A set of numbers is said to satisfy Benford's law if the leading digit d ($d \in \{1, \dots, 9\}$) occurs with **probability**

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\left(\frac{d+1}{d}\right) = \log_{10}\left(1 + \frac{1}{d}\right).$$

Numerically, the leading digits have the following distribution in Benford's law, where d is the leading digit and $P(d)$ the probability:

The quantity $P(d)$ is proportional to the space between d and $d+1$ on a **logarithmic scale**. Therefore, this is the distribution expected if the **mantissae** of the **logarithms** of the numbers (but not the numbers themselves) are **uniformly and randomly distributed**. For example, a number x , constrained to lie between 1 and 10, starts with the digit 1 if $1 \leq x < 2$, and starts with the digit 9 if $9 \leq x < 10$. Therefore, x starts with the digit 1 if $\log 1 \leq \log x < \log 2$, or starts with 9 if $\log 9 \leq \log x < \log 10$. The interval $[\log 1, \log 2]$ is much wider than the interval $[\log 9, \log 10]$ (0.30 and 0.05 respectively); therefore if $\log x$ is uniformly and randomly distributed, it is much more likely to fall into the wider interval than the narrower interval, i.e. more likely to start with 1 than with 9. The probabilities are proportional to the interval widths, and this gives the equation above. (The above discussion assumed x is between 1 and 10, but the result is the same no matter how many digits x has before the decimal point.)

An extension of Benford's law predicts the distribution of first digits in other **bases** besides **decimal**; in fact, any base $b \geq 2$. The general form is:

$$P(d) = \log_b(d+1) - \log_b(d) = \log_b\left(1 + \frac{1}{d}\right).$$

For $b = 2$ (the **binary number system**), Benford's law is true but trivial: All binary numbers (except for 0) start with the digit 1. (On the other hand, the **generalization of Benford's law to second and later digits** is not trivial, even for binary numbers.) Also, Benford's law does not apply to unary systems such as tally marks.

1.10.2 Example

Examining a list of the heights of the 60 tallest structures in the world by category shows that 1 is by far the most common leading digit, *irrespective of the unit of measurement*:

1.10.3 History

The discovery of Benford's law goes back to 1881, when the American astronomer Simon Newcomb noticed that in logarithm tables the earlier pages (that started with 1) were much more worn than the other pages.^[5] Newcomb's published result is the first known instance of this observation and includes a distribution on the second digit, as well. Newcomb proposed a law that the probability of a single number N being the first digit of a number was equal to $\log(N + 1) - \log(N)$.

The phenomenon was again noted in 1938 by the physicist Frank Benford,^[4] who tested it on data from 20 different domains and was credited for it. His data set included the surface areas of 335 rivers, the sizes of 3259 US populations, 104 physical constants, 1800 molecular weights, 5000 entries from a mathematical handbook, 308 numbers contained in an issue of *Reader's Digest*, the street addresses of the first 342 persons listed in *American Men of Science* and 418 death rates. The total number of observations used in the paper was 20,229. This discovery was later named after Benford (making it an example of Stigler's Law).

In 1995, Ted Hill proved the result about mixed distributions mentioned below.^[6]

1.10.4 Explanations

Arno Berger and Ted Hill have stated that, "The widely known phenomenon called Benford's law continues to defy attempts at an easy derivation".^[1]

However, limited explanations of Benford's law have been offered.

Overview

Benford's law states that the fractional part of the logarithm of the data is uniformly distributed between 0 and 1. It tends to apply most accurately to data that are distributed uniformly across many orders of magnitude. As a rule, the more orders of magnitude that the data evenly covers, the more accurately Benford's law applies.

For instance, one can expect that Benford's law would apply to a list of numbers representing the populations of UK villages, or representing the values of small insurance claims. But if a "village" is defined as a settlement with population between 300 and 999, or a "small insurance claim" is defined as a claim between \$50 and \$99, then Benford's law will not apply.^{[7][8]}

Consider the probability distributions shown below, referenced to a log scale.^[9] In each case, the total area in red is the relative probability that the first digit is 1, and the total area in blue is the relative probability that the first digit is 8.

For the left distribution, the size of the areas of red and blue are approximately proportional to the widths of each red and blue bar. Therefore the numbers drawn from this distribution will approximately follow Benford's law. On the other hand, for the right distribution, the ratio of the areas of red and blue is very different from the ratio of the widths of each red and blue bar. Rather, the relative areas of red and blue are determined more by the height of the bars than the widths. Accordingly, the first digits in this distribution do not satisfy Benford's law at all.^[8]

Thus, real-world distributions that span several orders of magnitude rather uniformly (e.g. populations of villages / towns / cities, stock-market prices), are likely to satisfy Benford's law to a very high accuracy. On the other hand, a distribution that is mostly or entirely within one order of magnitude (e.g. heights of human adults, or IQ scores) is unlikely to satisfy Benford's law very accurately, or at all.^{[7][8]} However, it is not a sharp line: As the distribution gets narrower, the discrepancies from Benford's law typically increase gradually.

In terms of conventional probability density (referenced to a linear scale rather than log scale, i.e. $P(x)dx$ rather than $P(\log x) d(\log x)$), the equivalent criterion is that Benford's law will be very accurately satisfied when $P(x)$ is approximately proportional to $1/x$ over several orders-of-magnitude variation in x .^[9]

This discussion is not a *full* explanation of Benford's law, because we have not explained *why* we so often come across data-sets that, when plotted as a probability distribution of the logarithm of the variable, are relatively uniform over several orders of magnitude.^[10] The following sections give examples of how this might happen.

Outcomes of exponential growth processes

Here is a simple example where Benford's law would occur. 1000 cells of bacteria are introduced into a dish full of food. The number of bacteria **grows exponentially**, doubling each day. Every few hours for 30 days, one counts the number of bacteria that are in the dish, and writes down that number on a list. (By the end of 30 days, there will be a trillion bacteria.) Then this list of numbers will follow Benford's law quite accurately.

Why? Remember, the number of bacteria is growing exponentially, doubling each day. On the first day, the number of bacteria is increasing from 1000 towards 2000: The first digit is 1 the whole day. On the second day, there are 2000 bacteria increasing towards 4000: The first digit is 2 for fourteen hours and 3 for six hours. On the third day, there are 4000 bacteria increasing towards 8000: The first digit will pass through 4, 5, 6, and 7, spending less and less time in each digit. The next day, there are 8000 bacteria increasing towards 16,000. The leading digit will pass rapidly through 8 and 9 in a few hours, but then once there are 10,000 bacteria, the first digit will be 1 for a whole 24 hours, until the number of bacteria gets to 20,000.

From this example, it can be seen that the first digit is 1 with the highest probability, and 9 with the lowest.

Another way to think about it is: An exponentially-growing quantity is moving rightward on a log-scale at a constant rate. If we measure the number of bacteria at a random time in the 30-day window, we will get a random point on the log-scale, uniformly distributed in that corresponding window (about 6 orders of magnitude). As explained in the previous section, we expect this kind of probability distribution to satisfy Benford's law with high accuracy.

This example makes it plausible that data tables that involve measurements of exponentially growing quantities will agree with Benford's Law. But the law also describes many data-sets which do not have any apparent relation to exponential growth.

Multiplicative fluctuations

Many real-world examples of Benford's law arise from multiplicative fluctuations.^[11] For example, if a stock price starts at \$100, and then each day it gets multiplied by a randomly-chosen factor between 0.99 and 1.01, then over an extended period of time the probability distribution of its price satisfies Benford's law with higher and higher accuracy.

The reason is that the *logarithm* of the stock price is undergoing a **random walk**, so over time its probability distribution will get more and more broad and uniform (see [above](#)).^[11] (More technically, the **central limit theorem** says that multiplying more and more random variables will create a **log-normal distribution** with larger and larger variance, so eventually it covers many orders of magnitude almost uniformly.)

Unlike multiplicative fluctuations, *additive* fluctuations do not lead to Benford's law: They lead instead to **normal probability distributions** (again by the **central limit theorem**), which do not satisfy Benford's law. For example, the "number of heartbeats that I experience on a given day" can be written as the *sum* of many random variables (e.g. the sum of heartbeats per minute over all the minutes of the day), so this quantity is *unlikely* to follow Benford's law. By contrast, that hypothetical stock price described above can be written as the *product* of many random variables (i.e. the price change factor for each day), so is *likely* to follow Benford's law quite well.

Scale invariance

If there is a list of lengths, the distribution of first digits of numbers in the list may be generally similar regardless of whether all the lengths are expressed in metres, or yards, or feet, or inches, etc.

This is not *always* the case. For example, the height of adult humans almost always starts with a 1 or 2 when measured in meters, and almost always starts with 4, 5, 6, or 7 when measured in feet.

But consider a list of lengths that is spread evenly over many orders of magnitude. For example, a list of 1000 lengths

mentioned in scientific papers will include the measurements of molecules, bacteria, plants, and galaxies. If one writes all those lengths in meters, or writes them all in feet, it is reasonable to expect that the distribution of first digits should be the same on the two lists.

In these situations, where the distribution of first digits of a data set is **scale invariant** (or independent of the units that the data are expressed in), the distribution of first digits is always given by Benford's Law.^{[12][13]} To be sure of approximate agreement with Benford's Law, the data has to be approximately invariant when scaled up by any factor up to 10. A **lognormally** distributed data set with wide dispersion has this approximate property, as do some of the examples mentioned above.

For example, the first (non-zero) digit on this list of lengths should have the same distribution whether the unit of measurement is feet or yards. But there are three feet in a yard, so the probability that the first digit of a length in yards is 1 must be the same as the probability that the first digit of a length in feet is 3, 4, or 5. Applying this to all possible measurement scales gives the logarithmic distribution of Benford's law.

Multiple probability distributions

For numbers drawn from certain distributions (IQ scores, human heights) the Law fails to hold because these variates obey a normal distribution which is known not to satisfy Benford's law,^[14] since normal distributions can't span several orders of magnitude and the mantissae of their logarithms will not be (even approximately) uniformly distributed.

However, if one "mixes" numbers from those distributions, for example by taking numbers from newspaper articles, Benford's law reappears. This can also be proven mathematically: if one repeatedly "randomly" chooses a **probability distribution** (from an uncorrelated set) and then randomly chooses a number according to that distribution, the resulting list of numbers will obey Benford's Law.^{[6][15]} A similar probabilistic explanation for the appearance of Benford's Law in everyday-life numbers has been advanced by showing that it arises naturally when one considers mixtures of uniform distributions.^[16]

1.10.5 Applications

Accounting fraud detection

In 1972, **Hal Varian** suggested that the law could be used to detect possible fraud in lists of socio-economic data submitted in support of public planning decisions. Based on the plausible assumption that people who make up figures tend to distribute their digits fairly uniformly, a simple comparison of first-digit frequency distribution from the data with the expected distribution according to Benford's Law ought to show up any anomalous results.^[17] Following this idea, **Mark Nigrini** showed that Benford's Law could be used in **forensic accounting** and **auditing** as an indicator of accounting and expenses fraud.^[18] In practice, applications of Benford's Law for fraud detection routinely use more than the first digit.^[18]

Legal status

In the United States, evidence based on Benford's law has been admitted in criminal cases at the federal, state, and local levels.^[19]

Election data

Benford's Law has been invoked as evidence of fraud in the **2009 Iranian elections**,^[20] and also used to analyze other election results. However, other experts consider Benford's Law essentially useless as a statistical indicator of election fraud in general.^{[21][22]}

Macroeconomic data

Similarly, the macroeconomic data the Greek government reported to the European Union before entering the **eurozone** was shown to be probably fraudulent using Benford's law, albeit years after the country joined.^[23]

Genome data

The number of **open reading frames** and their relationship to genome size differs between **eukaryotes** and **prokaryotes** with the former showing a log-linear relationship and the latter a linear relationship. Benford's law has been used to test this observation with an excellent fit to the data in both cases.^[24]

Scientific fraud detection

A test of regression coefficients in published papers showed agreement with Benford's law.^[25] As a comparison group subjects were asked to fabricate statistical estimates. The fabricated results failed to obey Benford's law.

1.10.6 Statistical tests

Although the **chi squared** test has been used to test for compliance with Benford's law it has low statistical power when used with small samples.

The **Kolmogorov–Smirnov test** and the **Kuiper test** are more powerful when the sample size is small particularly when Stephens's corrective factor is used.^[26] These tests may be overly conservative when applied to discrete distribution. Values for the Benford test have been generated by Morrow.^[27] The critical values of the test statistics are shown below:

Two alternative tests specific to this law have been published: first, the max (m) statistic^[28] is given by

$$m = \sqrt{N} \cdot \max_{i=1}^9 \left\{ \Pr(X_{\text{FSD}} \text{ has } = i) - \log_{10}(1 + 1/i) \right\},$$

and secondly, the distance (d) statistic^[29] is given by

$$d = \sqrt{N \cdot \sum_{i=1}^9 \left[\Pr(X_{\text{FSD}} \text{ has } = i) - \log_{10}(1 + 1/i) \right]^2},$$

where FSD is the First Significant Digit and N is the sample size. Morrow has determined the critical values for both these statistics, which are shown below:^[27]

Nigrini^[30] has suggested the use of a **z statistic**

$$z = \frac{|p_o - p_e| - \frac{1}{2n}}{s_i}$$

with

$$s_i = \left[\frac{p_e(1 - p_e)}{n} \right]^{1/2},$$

where $|x|$ is the absolute value of x , n is the sample size, $1/(2n)$ is a continuity correction factor, p_e is the proportion expected from Benford's law and p_o is the observed proportion in the sample.

Morrow has also shown that for any random variable X (with a continuous pdf) divided by its standard deviation (σ), a value A can be found such that the probability of the distribution of the first significant digit of the random variable (X / σ)^A will differ from Benford's Law by less than $\varepsilon > 0$.^[27] The value of A depends on the value of ε and the distribution of the random variable.

A method of accounting fraud detection based on bootstrapping and regression has been proposed.^[31]

1.10.7 Generalization to digits beyond the first

It is possible to extend the law to digits beyond the first.^[32] In particular, the probability of encountering a number starting with the string of digits n is given by:

$$\log_{10}(n+1) - \log_{10}(n) = \log_{10}\left(1 + \frac{1}{n}\right)$$

(For example, the probability that a number starts with the digits 3, 1, 4 is $\log_{10}(1 + 1/314) \approx 0.0014$.) This result can be used to find the probability that a particular digit occurs at a given position within a number. For instance, the probability that a "2" is encountered as the second digit is^[32]

$$\log_{10}\left(1 + \frac{1}{12}\right) + \log_{10}\left(1 + \frac{1}{22}\right) + \cdots + \log_{10}\left(1 + \frac{1}{92}\right) \approx 0.109$$

And the probability that d ($d = 0, 1, \dots, 9$) is encountered as the n -th ($n > 1$) digit is

$$\sum_{k=10^{n-2}}^{10^{n-1}-1} \log_{10}\left(1 + \frac{1}{10k+d}\right)$$

The distribution of the n -th digit, as n increases, rapidly approaches a uniform distribution with 10% for each of the ten digits.^[32] Four digits is often enough to assume a uniform distribution of 10% as '0' appears 10.0176% of the time in the fourth digit while '9' appears 9.9824% of the time.

1.10.8 Tests with common distributions

Benford's law was empirically tested against the numbers (up to the 10th digit) generated by a number of important distributions, including the uniform distribution, the exponential distribution, the half-normal distribution, the right-truncated normal, the normal distribution, the chi square distribution and the log normal distribution.^[14] In addition to these the ratio distribution of two uniform distributions, the ratio distribution of two exponential distributions, the ratio distribution of two half-normal distributions, the ratio distribution of two right-truncated normal distributions, the ratio distribution of two chi-square distributions (the F distribution) and the log normal distribution were tested.

The uniform distribution as might be expected does not obey Benford's law. In contrast, the ratio distribution of two uniform distributions is well described by Benford's law. Benford's law also describes the exponential distribution and the ratio distribution of two exponential distributions well. Although the half-normal distribution does not obey Benford's law, the ratio distribution of two half-normal distributions does. Neither the right-truncated normal distribution nor the ratio distribution of two right-truncated normal distributions are well described by Benford's law. This is not surprising as this distribution is weighted towards larger numbers. Neither the normal distribution nor the ratio distribution of two normal distributions (the Cauchy distribution) obey Benford's law. The fit of chi square distribution depends on the degrees of freedom (df) with good agreement with $df = 1$ and decreasing agreement as the df increases. The F distribution is fitted well for low degrees of freedom. With increasing dfs the fit decreases but much more slowly than the chi square distribution. The fit of the log-normal distribution depends on the mean and the variance of the distribution. The variance

has a much greater effect on the fit than does the mean. Larger values of both parameters result in better agreement with the law. The ratio of two log normal distributions is a log normal so this distribution was not examined.

Other distributions that have been examined include the Muth distribution, Gompertz distribution, Weibull distribution, gamma distribution, log-logistic distribution and the exponential power distribution all of which show reasonable agreement with the law.^{[28][33]} The Gumbel distribution – a density increases with increasing value of the random variable – does not show agreement with this law.^[33]

1.10.9 Distributions known to obey Benford's law

Some well-known infinite integer sequences provably satisfy Benford's Law exactly (in the asymptotic limit as more and more terms of the sequence are included). Among these are the Fibonacci numbers,^{[34][35]} the factorials,^[36] the powers of 2,^{[37][38]} and the powers of almost any other number.^[37]

Likewise, some continuous processes satisfy Benford's Law exactly (in the asymptotic limit as the process continues through time). One is an exponential growth or decay process: If a quantity is exponentially increasing or decreasing in time, then the percentage of time that it has each first digit satisfies Benford's Law asymptotically (i.e. increasing accuracy as the process continues through time).

1.10.10 Distributions known to not obey Benford's law

Square roots and reciprocals do not obey this law.^[39] The 1974 Vancouver, Canada telephone book violates Benford's law because regulations require that telephone numbers have a fixed number of digits and do not begin with 1. Benford's law is violated by the populations of all places with population at least 2500 from five US states according to the 1960 and 1970 censuses, where only 19% began with digit 1 but 20% began with digit 2, for the simple reason that the truncation at 2500 introduces statistical bias.^[39] The terminal digits in pathology reports violate Benford's law due to rounding, and the fact that terminal digits are never expected to follow Benford's law in the first place.^[40]

1.10.11 Criteria for distributions expected and not expected to obey Benford's Law

A number of criteria—applicable particularly to accounting data—have been suggested where Benford's Law can be expected to apply and not to apply.^[41]

Distributions that can be expected to obey Benford's Law

- When the mean is greater than the median and the skew is positive
- Numbers that result from mathematical combination of numbers: e.g. quantity $\times$ price
- Transaction level data: e.g. disbursements, sales
- Numbers produced when doing any multiplicative calculations with an Oughtred slide rule, since the answers naturally fall into the right logarithmic distribution.

Distributions that would not be expected to obey Benford's Law

- Where numbers are assigned sequentially: e.g. check numbers, invoice numbers
- Where numbers are influenced by human thought: e.g. prices set by psychological thresholds (\$1.99)
- Accounts with a large number of firm-specific numbers: e.g. accounts set up to record \$100 refunds
- Accounts with a built-in minimum or maximum
- Where no transaction is recorded

1.10.12 Moments

Moments of random variables for the digits 1 to 9 following this law have been calculated:^[42]

- mean 3.440
- variance 6.057
- skewness 0.796
- kurtosis −0.548

For the first and second digit distribution these values are also known:^[43]

- mean 38.590
- variance 621.832
- skewness 0.772
- kurtosis −0.547

A table of the exact probabilities for the joint occurrence of the first two digits according to Benford's law is available,^[43] as is the population correlation between the first and second digits:^[43] $\rho = 0.0561$.

1.10.13 See also

- Fraud detection in predictive analytics
- Zipf's law

1.10.14 References

- [1] Arno Berger and Theodore P Hill, Benford's Law Strikes Back: No Simple Explanation in Sight for Mathematical Gem, 2011
- [2] Weisstein, Eric W. "Benford's Law". *MathWorld*, A Wolfram web resource. Retrieved 7 June 2015.
- [3] Paul H. Kvam, Brani Vidakovic, *Nonparametric Statistics with Applications to Science and Engineering*, p. 158
- [4] Frank Benford (March 1938). "The law of anomalous numbers". *Proc. Am. Philos. Soc.* **78** (4): 551–572. JSTOR 984802. (subscription required)
- [5] Simon Newcomb (1881). "Note on the frequency of use of the different digits in natural numbers". *American Journal of Mathematics* (American Journal of Mathematics, Vol. 4, No. 1) **4** (1/4): 39–40. doi:10.2307/2369148. JSTOR 2369148. (subscription required)
- [6] Theodore P. Hill (1995). "A Statistical Derivation of the Significant-Digit Law" (PDF). *Statistical Science* **10**: 354–363. doi:10.1214/ss/1177009869. MR 1421567.
- [7] Steven W. Smith. "The Scientist and Engineer's Guide to Digital Signal Processing, chapter 34, Explaining Benford's Law". Retrieved 15 December 2012. (especially section 10).
- [8] Fewster, R. M. (2009). "A simple explanation of Benford's Law". *The American Statistician* **63** (1): 26–32. doi:10.1198/tast.2009.0005.
- [9] This section discusses and plots probability distributions of the logarithms of a variable. This is not the same as taking a regular probability distribution of a variable, and simply plotting it on a log scale. Instead, one multiplies the distribution by a certain function. The log scale distorts the horizontal distances, so the height has to be changed also, in order for the area under each section of the curve to remain true to the original distribution. See, for example, . Specifically: $P(\log x)d(\log x) = (1/x)P(x)dx$.

- [10] Arno Berger and Theodore P Hill, *Benford's Law Strikes Back: No Simple Explanation in Sight for Mathematical Gem*, 2011. The authors describe this argument, but say it “still leaves open the question of why it is reasonable to assume that the logarithm of the spread, as opposed to the spread itself—or, say, the log log spread—should be large.” Moreover, they say: “assuming large spread on a logarithmic scale is *equivalent* to assuming an approximate conformance with [Benford's law]” (*italics added*), something which they say lacks a “simple explanation”.
- [11] L. Pietronero, E. Tosatti, V. Tosatti, A. Vespignani (2001). “Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf”. *Physica A* **293**: 297–304. Bibcode:2001PhyA..293..297P. doi:10.1016/S0378-4371(00)00633-6.
- [12] Roger S. Pinkham, *On the Distribution of First Significant Digits*, Ann. Math. Statist. Volume 32, Number 4 (1961), 1223-1230.
- [13] MathWorld – Benford's Law
- [14] Formann, A. K. (2010). Morris, Richard James, ed. “The Newcomb-Benford Law in Its Relation to Some Common Distributions”. *PLoS ONE* **5** (5): e10541. Bibcode:2010PLoSO...510541F. doi:10.1371/journal.pone.0010541. PMC 2866333. PMID 20479878.
- [15] Theodore P. Hill (July–August 1998). “The first digit phenomenon” (PDF). *American Scientist* **86** (4): 358. Bibcode:1998AmSci..86..358H. doi:10.1511/1998.4.358.
- [16] Élise Janvresse and Thierry de la Rue (2004), “From Uniform Distributions to Benford's Law”, *Journal of Applied Probability*, 41 1203–1210 doi:10.1239/jap/1101840566 MR 2122815 preprint
- [17] Varian, Hal (1972). “Benford's Law (Letters to the Editor)”. *The American Statistician* **26** (3): 65. doi:10.1080/00031305.1972.10478934.
- [18] Nigrini, Mark J. (May 1999). “I've Got Your Number:How a mathematical phenomenon can help CPAs uncover fraud and other irregularities”. *Journal of Accountancy*.
- [19] “From Benford to Erdős”. *Radio Lab*. Episode 2009-10-09. 2009-09-30.
- [20] Stephen Battersby Statistics hint at fraud in Iranian election *New Scientist* 24 June 2009
- [21] Joseph Deckert, Mikhail Myagkov and Peter C. Ordeshook, (2010) *The Irrelevance of Benford's Law for Detecting Fraud in Elections*, Caltech/MIT Voting Technology Project Working Paper No. 9
- [22] Charles R. Tolle, Joanne L. Budzien, and Randall A. LaViolette (2000) *Do dynamical systems follow Benford's Law?*, Chaos 10, 2, pp.331–336 (2000); doi:10.1063/1.166498
- [23] Müller, Hans Christian: *Greece Was Lying About Its Budget Numbers*. *Forbes*. 12 September 2011.
- [24] Friar, JL; Goldman, T; Pérez-Mercader, J (2012). “Genome sizes and the benford distribution”. *PLOS ONE* **7** (5): e36624. arXiv:1205.6512. Bibcode:2012PLoSO...736624F. doi:10.1371/journal.pone.0036624.
- [25] Diekmann A (2007) Not the First Digit! Using Benford's Law to detect fraudulent scientific data. *J Appl Stat* 34 (3) 321–329, doi:10.1080/02664760601004940
- [26] Stephens, M. A. (1970). “Use of the Kolmogorov–Smirnov, Cramér–Von Mises and Related Statistics without Extensive Tables”. *Journal of the Royal Statistical Society, Series B* **32** (1): 115–122. Retrieved 2013-03-09.
- [27] Morrow, J. (2010) “Benford's Law, Families of Distributions and a test basis”, UW-Madison
- [28] Leemis, L. M.; Schmeiser, B. W.; Evans, D. L. (2000). “Survival distributions satisfying Benford's Law”. *The American Statistician* **54** (4): 236–241. doi:10.1080/00031305.2000.10474554.
- [29] Cho, W. K. T.; Gaines, B. J. (2007). “Breaking the (Benford) law: Statistical fraud detection in campaign finance”. *The American Statistician* **61** (3): 218–223. doi:10.1198/000313007X223496.
- [30] Nigrini, M. (1996). “A taxpayer compliance application of Benford's Law”. *J Amer Tax Assoc* **18**: 72–91.
- [31] Suh, I. S.; Headrick, T. C.; Minaburo, S. (2011). “An effective and efficient analytic technique: A bootstrap regression procedure and Benford's Law”. *J Forensic & Investigative Accounting* **3** (3).
- [32] Theodore P. Hill, “The Significant-Digit Phenomenon”, *The American Mathematical Monthly*, Vol. 102, No. 4, (Apr., 1995), pp. 322–327. Official web link (subscription required). Alternate, free web link.

- [33] Dümmbgen, L; Leuenberger, C (2008). "Explicit bounds for the approximation error in Benford's Law". *Elect Comm in Probab* **13**: 99–112. doi:10.1214/ECP.v13-1358.
- [34] L. C. Washington, "Benford's Law for Fibonacci and Lucas Numbers", *The Fibonacci Quarterly*, **19.2**, (1981), 175–177
- [35] Duncan, R. L. (1967). "An Application of Uniform Distribution to the Fibonacci Numbers". *The Fibonacci Quarterly* **5**: 137–140.
- [36] Sarkar, P. B. (1973). "An Observation on the Significant Digits of Binomial Coefficients and Factorials", *Sankhya*. B, **35**: 363–364.
- [37] In general, the sequence k^1, k^2, k^3 , etc., satisfies Benford's Law exactly, under the condition that $\log_{10} k$ is an irrational number. This is a straightforward consequence of the equidistribution theorem.
- [38] That the first 100 powers of 2 approximately satisfy Benford's Law is mentioned by Ralph Raimi. Raimi, Ralph A. (1976). "The First Digit Problem". *American Mathematical Monthly* **83** (7): 521–538. doi:10.2307/2319349.
- [39] Raimi, RA (1976). "The first digit problem". *American Mathematical Monthly* **83**: 521–538. doi:10.2307/2319349.
- [40] Beer, TW (2009). "Terminal digit preference: beware of Benford's Law". *J Clin Pathol* **62**: 192. doi:10.1136/jcp.2008.061721.
- [41] Durtschi, C; Hillison, W; Pacini, C (2004). "The effective use of Benford's Law to assist in detecting fraud in accounting data". *J Forensic Accounting* **5**: 17–34.
- [42] Scott, P.D.; Fasli, M. (2001) "Benford's Law: An empirical investigation and a novel explanation". *CSM Technical Report* 349, Department of Computer Science, Univ. Essex
- [43] Suh, I.S.; Headrick, T.C. (2010). "A comparative analysis of the bootstrap versus traditional statistical procedures applied to digital analysis based on Benford's Law" (PDF). *Journal of Forensic and Investigative Accounting* **2** (2): 144–175.

1.10.15 Further reading

- Arno Berger and Theodore P. Hill (2015). *An Introduction to Benford's Law*. Princeton University Press. ISBN 978-0-691-16306-2.
- Alex Ely Kossovsky. *Benford's Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications*, 2014, World Scientific Publishing. ISBN 978-981-4583-68-8.
- "Benford's Law – from Wolfram MathWorld". Mathworld.wolfram.com. 14 June 2012. Retrieved 2012-06-26.
- Mark J. Nigrini (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. John Wiley & Sons. p. 330. ISBN 978-1-118-15285-0.
- Alessandro Gambini, Giovanni Mingari Scarpello, Daniele Ritelli; et al. (2012). "Probability of digits by dividing random numbers: A ψ and ζ functions approach". *Expositiones Mathematicae* **30** (4): 223–238. doi:10.1016/j.exmath.2012.03.001.
- Sehity; Hoelzl, Erik; Kirchler, Erich (2005). "Price developments after a nominal shock: Benford's Law and psychological pricing after the euro introduction". *International Journal of Research in Marketing* **22** (4): 471–480. doi:10.1016/j.ijresmar.2005.09.002.
- Nicolas Gauvrit; Jean-Paul Delahaye (2011). "Scatter and regularity implies Benford's Law...and more". *Zenil: Randomness through computation: some answers, more questions*. ISBN 9814327751: 58–69. arXiv:0910.1359. Bibcode:2009arXiv0910.1359G. doi:10.1142/9789814327756_0004.
- Bernhard Rauch1, Max Götttsche, Gernot Brähler, Stefan Engel (August 2011). "Fact and Fiction in EU-Governmental Economic Data". *German Economic Review* **12** (3): 243–255. doi:10.1111/j.1468-0475.2011.00542.x.
- Wendy Cho and Brian Gaines (August 2007). "Breaking the (Benford) Law: statistical fraud detection in campaign finance". *The American Statistician* **61** (3): 218–223. doi:10.1198/000313007X223496.
- Geiringer, Hilda; Furlan, L. V. (1948). "The Law of Harmony in Statistics: An Investigation of the Metrical Interdependence of Social Phenomena. by L. V. Furlan". *Journal of the American Statistical Association* (American Statistical Association) **43** (242): 325–328. doi:10.2307/2280379. JSTOR 2280379.

1.10.16 External links

General audience

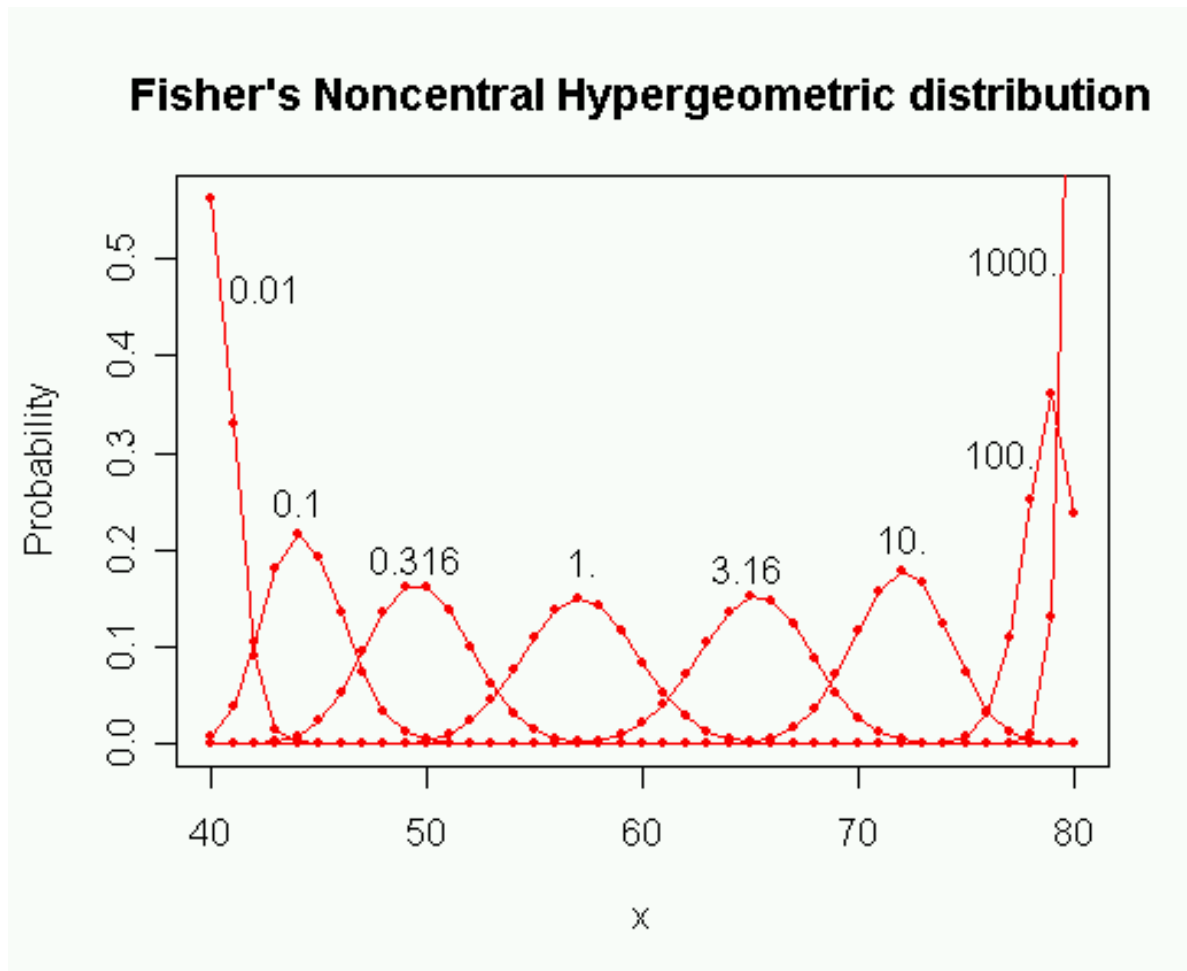
- [Benford Online Bibliography](#), an online bibliographic database on Benford's Law.
- [Companion website for Benford's Law by Mark Nigrini](#) Website includes 15 data sets, 10 Excel templates, photos, documents, and other miscellaneous items related to Benford's Law
- [Following Benford's Law, or Looking Out for No. 1](#), 1998 article from *The New York Times*.
- [A further five numbers: number 1 and Benford's law](#), BBC radio segment by Simon Singh
- [From Benford to Erdős](#), Radio segment from the Radiolab program
- [Looking out for number one](#) by Jon Walthoe, Robert Hunt and Mike Pearson, *Plus Magazine*, September 1999
- [Video showing Benford's Law applied to Web Data](#) (incl. Minnesota Lakes, US Census Data and Digg Statistics)
- [An illustration of Benford's Law](#), showing first-digit distributions of various sequences evolve over time, interactive.
- [Generate your own Benford numbers](#) A script for generating random numbers compliant with Benford's Law.
- [Testing Benford's Law](#) An open source project showing Benford's Law in action against publicly available datasets.
- [Testing Benford's Law in OLAP Cubes Implementation with Microsoft Analysis Services](#).
- [Mould, Steve. "Number 1 and Benford's Law". *Numberphile*. Brady Haran.](#)
- [A third of property values begin with a 1](#) An example of Benford's Law appearing in house price data.
- [Benford's Very Strange Law](#) - Professor John D. Barrow, lecture on Benford's Law.

More mathematical

- [Weisstein, Eric W., "Benford's Law", *MathWorld*.](#)
- [Benford's law, Zipf's law, and the Pareto distribution](#) by Terence Tao
- [Country Data and Benford's Law](#), Benford's Law from Ratios of Random Numbers at Wolfram Demonstrations Project.
- [Benford's Law Solved with Digital Signal Processing](#)
- [Interactive graphic: Univariate Distribution Relationships](#)



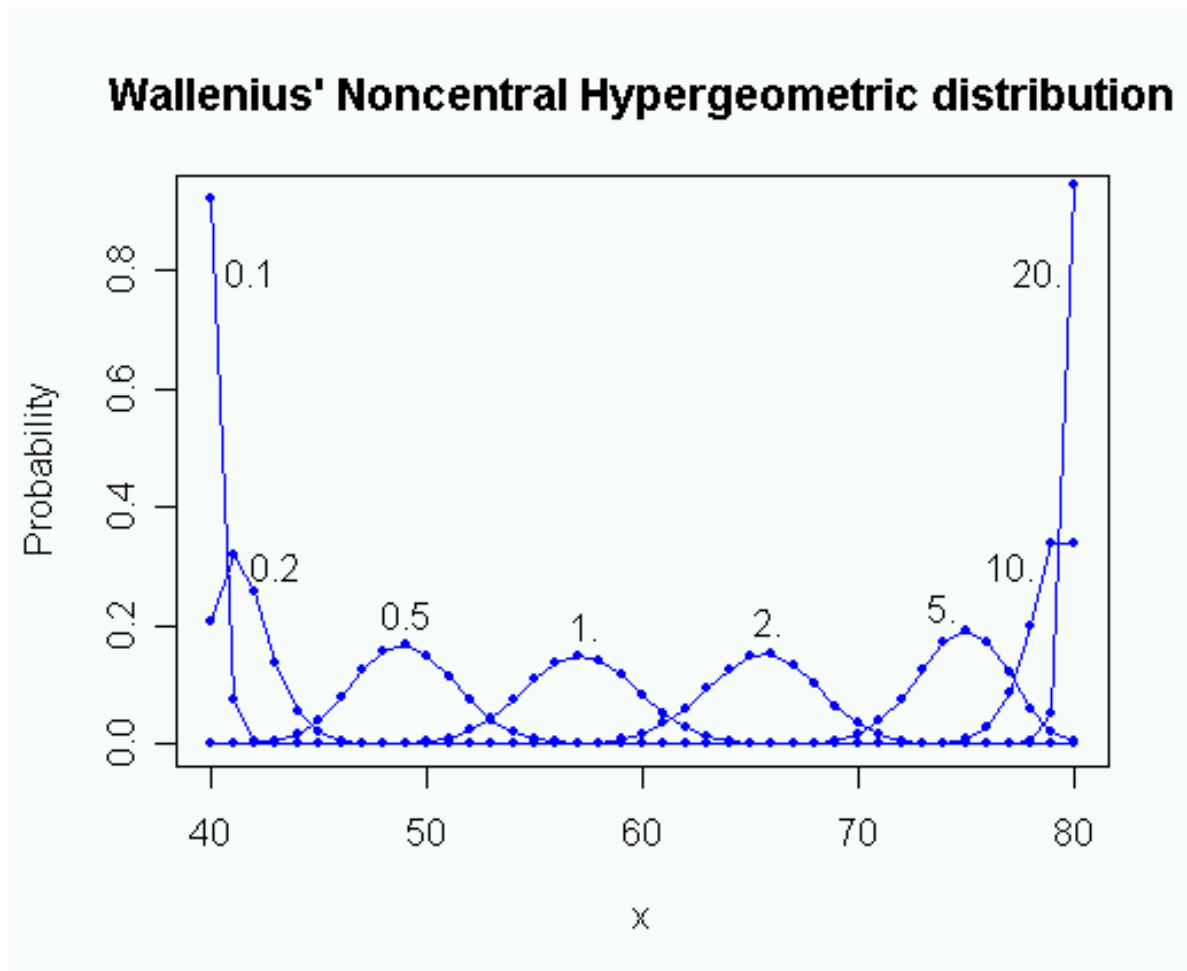
Biologist and statistician Ronald Fisher



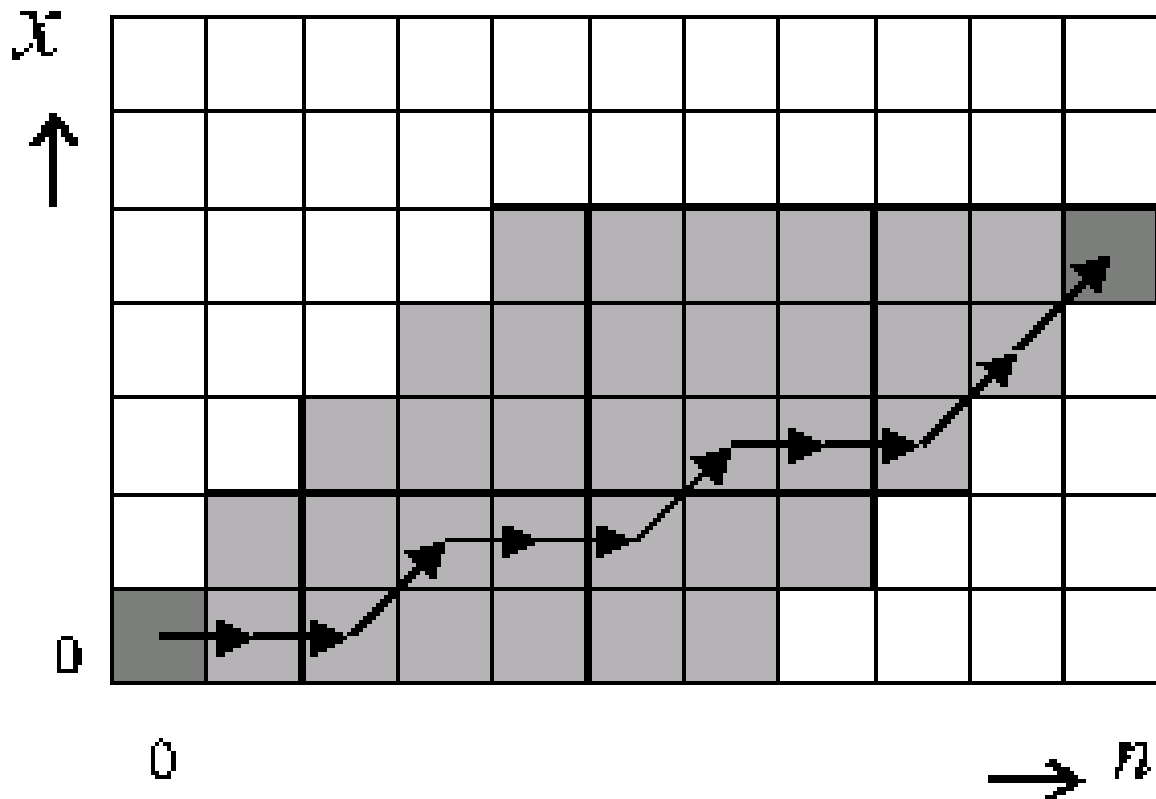
Probability mass function for Fisher's noncentral hypergeometric distribution for different values of the odds ratio ω .
 $m_1 = 80$, $m_2 = 60$, $n = 100$, $\omega = 0.01, \dots, 1000$



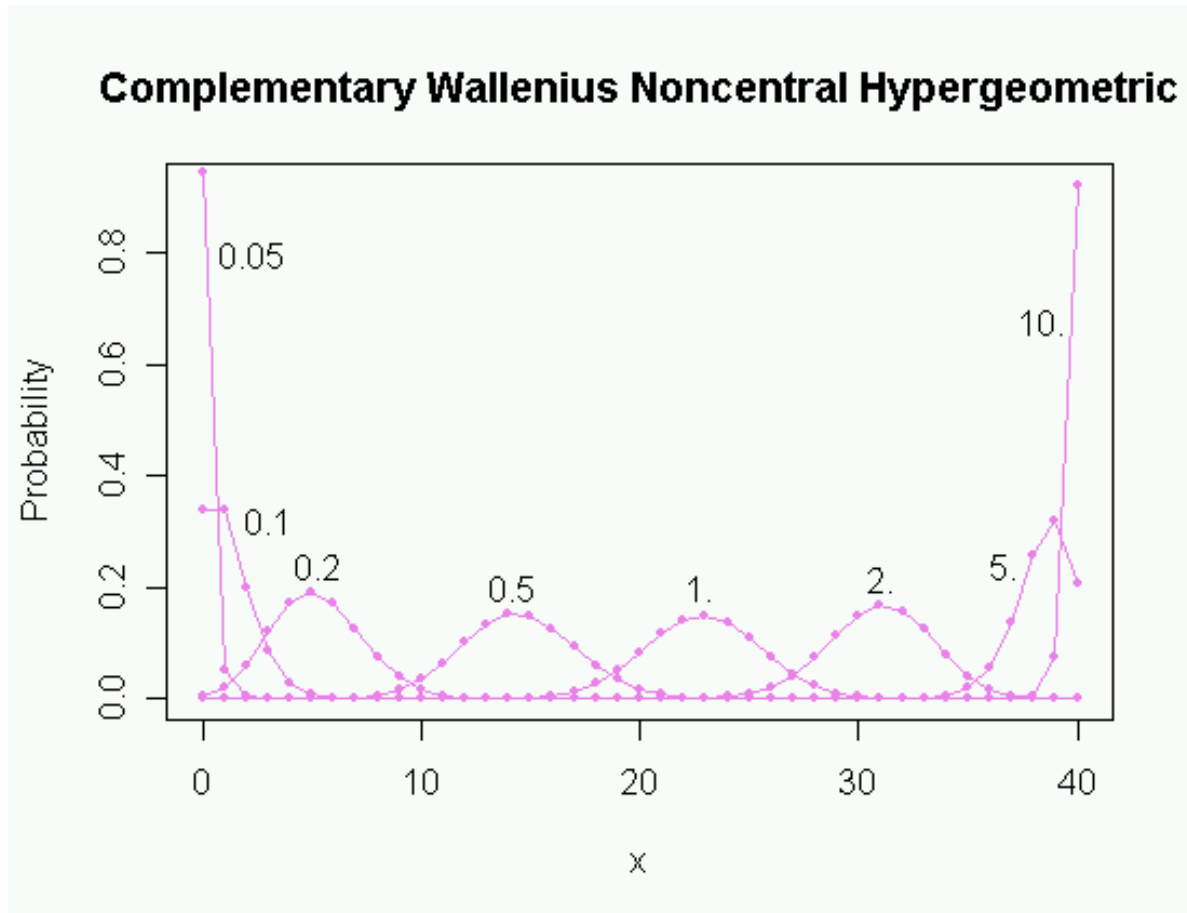
Biologist and statistician Ronald Fisher



Probability mass function for Wallenius' Noncentral Hypergeometric Distribution for different values of the odds ratio ω .
 $m_1 = 80, m_2 = 60, n = 100, \omega = 0.1 \dots 20$

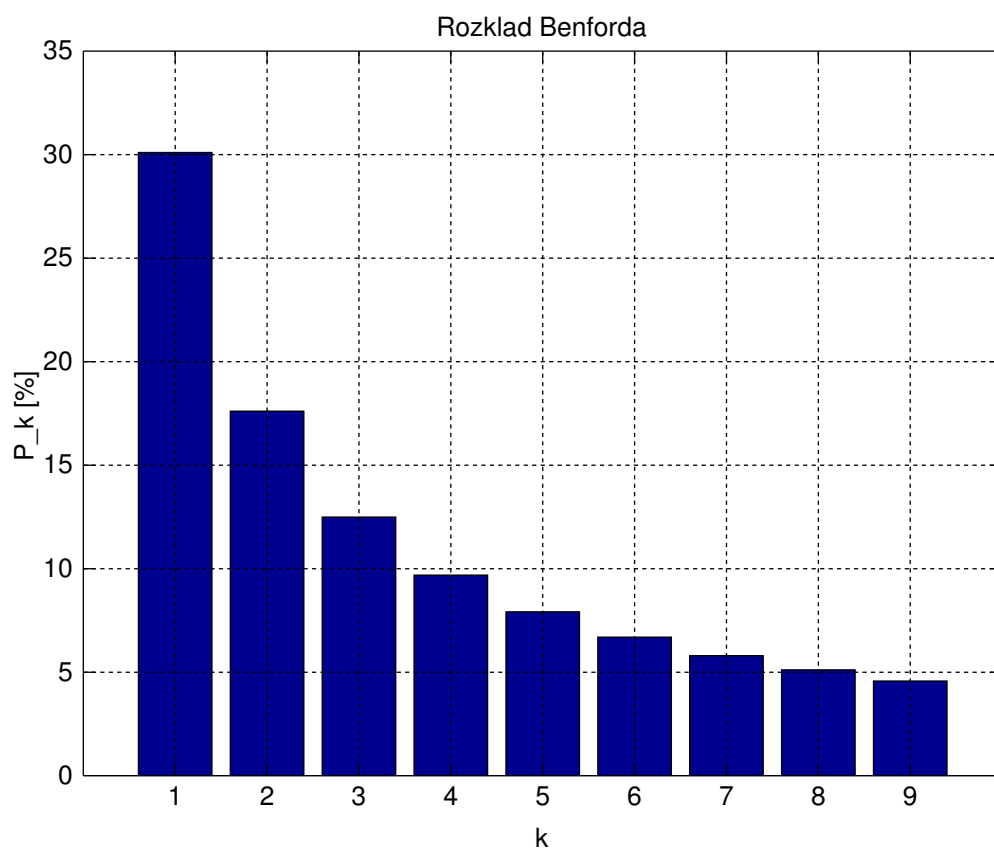


Recursive calculation of probability $f(x,n)$ in Wallenius' distribution. The light grey fields are possible points on the way to the final point. The arrows indicate an arbitrary trajectory.

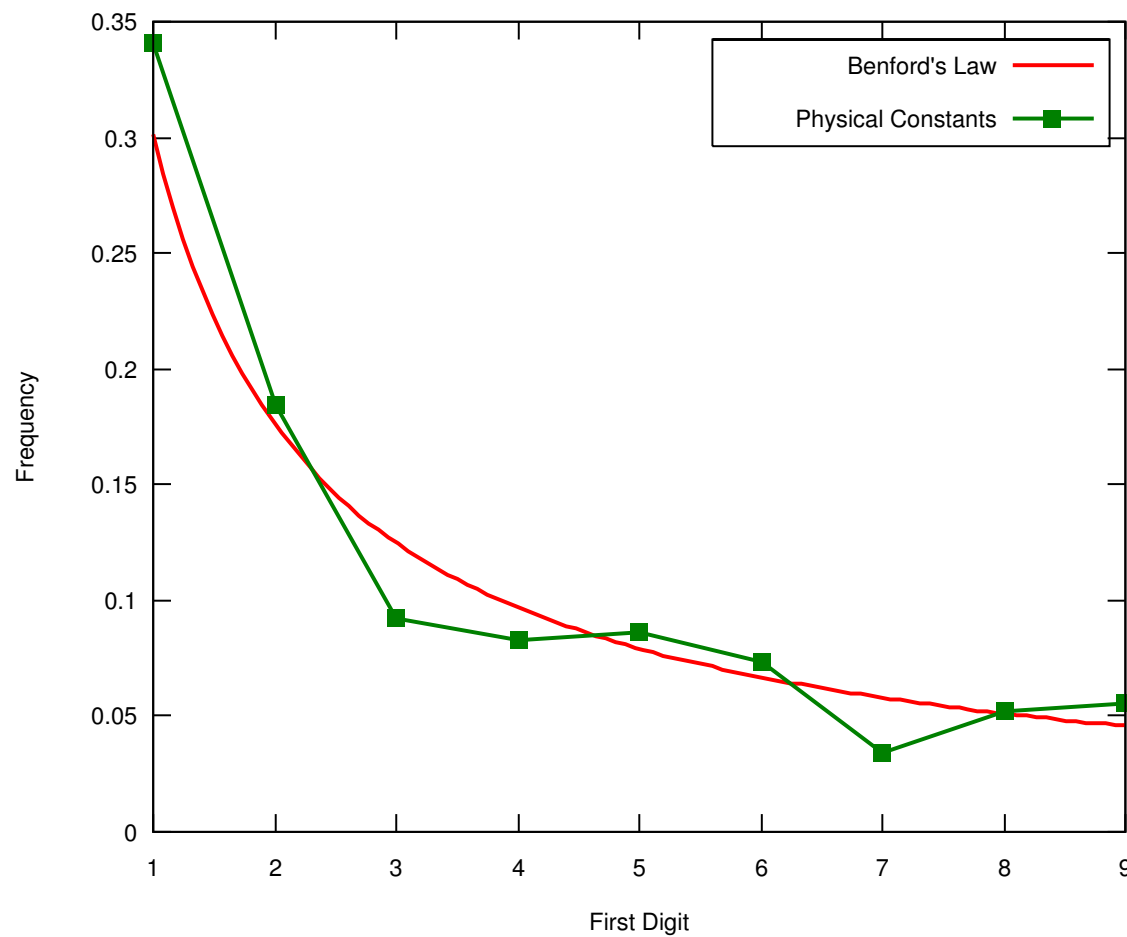


Probability mass function for the Complementary Wallenius' Noncentral Hypergeometric Distribution for different values of the odds ratio ω .

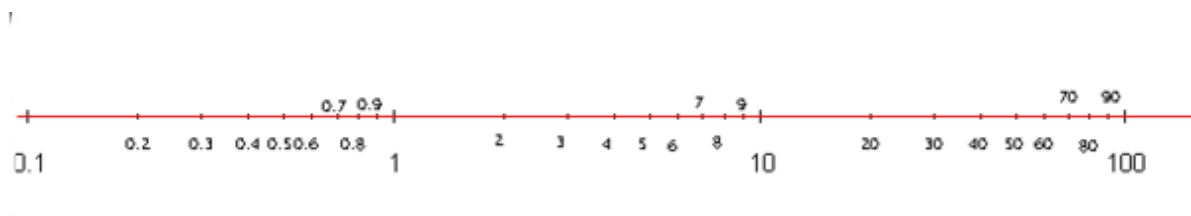
$m_1 = 80, m_2 = 60, n = 40, \omega = 0.05 \dots 10$



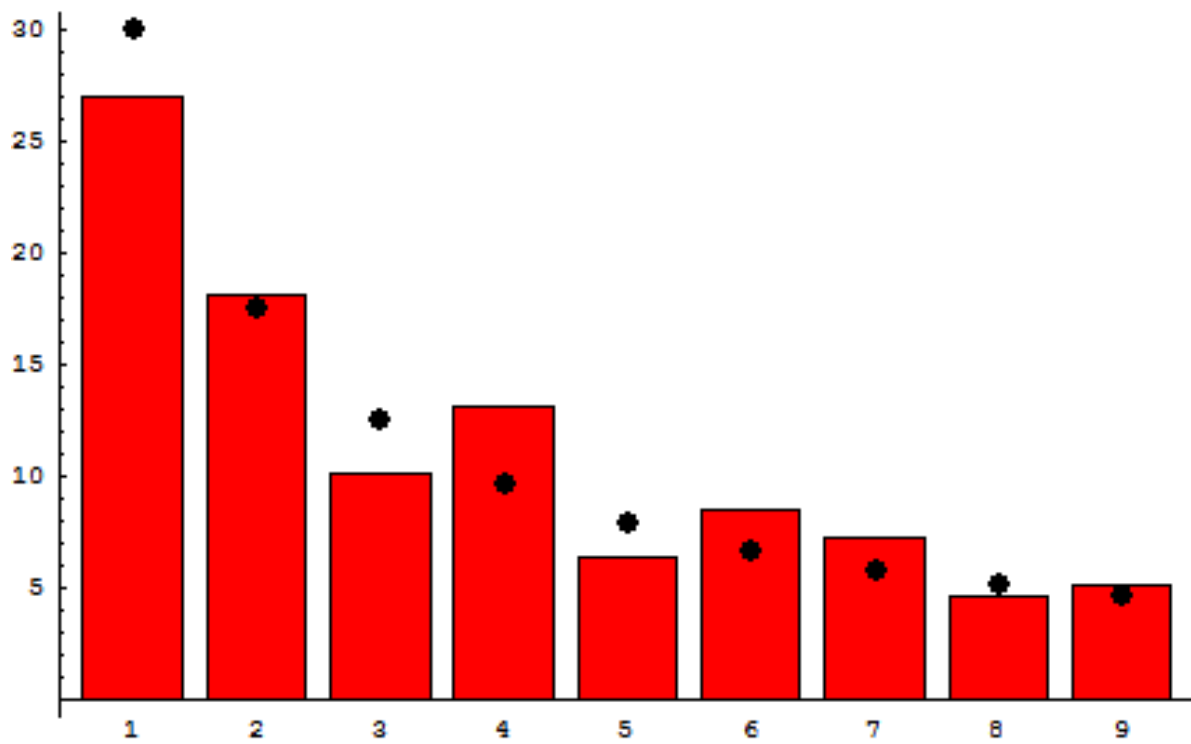
The distribution of first digits, according to Benford's law. Each bar represents a digit, and the height of the bar is the percentage of numbers that start with that digit.



Frequency of first significant digit of physical constants plotted against Benford's law



A logarithmic scale bar. Picking a random x position uniformly on this number line, roughly 30% of the time the first digit of the number will be 1.



Distribution of first digits (in %, red bars) in the population of the 237 countries of the world. Black dots indicate the distribution predicted by Benford's law.

Chapter 2

Continuous Distributions - Supported on semi-infinite intervals, usually $[0, \infty)$

2.1 Beta prime distribution

In probability theory and statistics, the **beta prime distribution** (also known as **inverted beta distribution** or **beta distribution of the second kind**^[1]) is an absolutely continuous probability distribution defined for $x > 0$ with two parameters α and β , having the probability density function:

$$f(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha, \beta)}$$

where B is a **Beta function**.

The cumulative distribution function is

$$F(x; \alpha, \beta) = I_{\frac{x}{1+x}}(\alpha, \beta),$$

where I is the **regularized incomplete beta function**.

The expectation value, variance, and other details of the distribution are given in the sidebox; for $\beta > 4$, the **excess kurtosis** is

$$\gamma_2 = 6 \frac{\alpha(\alpha + \beta - 1)(5\beta - 11) + (\beta - 1)^2(\beta - 2)}{\alpha(\alpha + \beta - 1)(\beta - 3)(\beta - 4)}$$

While the related **beta distribution** is the **conjugate prior distribution** of the parameter of a Bernoulli distribution expressed as a probability, the beta prime distribution is the conjugate prior distribution of the parameter of a Bernoulli distribution expressed in **odds**. The distribution is a **Pearson type VI** distribution.^[1]

The mode of a variate X distributed as $\beta'(\alpha, \beta)$ is $\hat{X} = \frac{\alpha-1}{\beta+1}$. Its mean is $\frac{\alpha}{\beta-1}$ if $\beta > 1$ (if $\beta \leq 1$ the mean is infinite, in other words it has no well defined mean) and its variance is $\frac{\alpha(\alpha+\beta-1)}{(\beta-2)(\beta-1)^2}$ if $\beta > 2$.

For $-\alpha < k < \beta$, the k -th moment $E[X^k]$ is given by

$$E[X^k] = \frac{B(\alpha + k, \beta - k)}{B(\alpha, \beta)}.$$

For $k \in \mathbb{N}$ with $k < \beta$, this simplifies to

$$E[X^k] = \prod_{i=1}^k \frac{\alpha + i - 1}{\beta - i}.$$

The cdf can also be written as

$$\frac{x^\alpha \cdot {}_2F_1(\alpha, \alpha + \beta, \alpha + 1, -x)}{\alpha \cdot B(\alpha, \beta)}$$

where ${}_2F_1$ is the Gauss's hypergeometric function ${}_2F_1$.

Differential equation

$$\left\{ (x^2 + x) f'(x) + f(x)(-\alpha + \beta x + x + 1) = 0, f(1) = \frac{2^{-\alpha-\beta}}{B(\alpha, \beta)} \right\}$$

2.1.1 Generalization

Two more parameters can be added to form the **generalized beta prime distribution**.

$p > 0$ shape (real)

$q > 0$ scale (real)

having the **probability density function**:

$$f(x; \alpha, \beta, p, q) = \frac{p \left(\frac{x}{q}\right)^{\alpha p - 1} \left(1 + \left(\frac{x}{q}\right)^p\right)^{-\alpha - \beta}}{q B(\alpha, \beta)}$$

with **mean**

$$\frac{q \Gamma(\alpha + \frac{1}{p}) \Gamma(\beta - \frac{1}{p})}{\Gamma(\alpha) \Gamma(\beta)} \text{ if } \beta p > 1$$

and **mode**

$$q \left(\frac{\alpha p - 1}{\beta p + 1} \right)^{\frac{1}{p}} \text{ if } \alpha p \geq 1$$

Note that if $p=q=1$ then the generalized beta prime distribution reduces to the **standard beta prime distribution**

Compound gamma distribution

The **compound gamma distribution**^[2] is the generalization of the beta prime when the scale parameter, q is added, but where $p=1$. It is so named because it is formed by **compounding** two **gamma distributions**:

$$\beta'(x; \alpha, \beta, 1, q) = \int_0^\infty G(x; \alpha, p) G(p; \beta, q) dp$$

where $G(x; a, b)$ is the gamma distribution with shape a and *inverse scale* b . This relationship can be used to generate random variables with a compound gamma, or beta prime distribution.

The mode, mean and variance of the compound gamma can be obtained by multiplying the mode and mean in the above infobox by q and the variance by q^2 .

2.1.2 Properties

- If $X \sim \beta'(\alpha, \beta)$ then $\frac{1}{X} \sim \beta'(\beta, \alpha)$.
- If $X \sim \beta'(\alpha, \beta, p, q)$ then $kX \sim \beta'(\alpha, \beta, p, kq)$.
- $\beta'(\alpha, \beta, 1, 1) = \beta'(\alpha, \beta)$

2.1.3 Related distributions

- If $X \sim F(\alpha, \beta)$ then $\frac{\alpha}{\beta} X \sim \beta'(\frac{\alpha}{2}, \frac{\beta}{2})$
- If $X \sim \text{Beta}(\alpha, \beta)$ then $\frac{X}{1-X} \sim \beta'(\alpha, \beta)$
- If $X \sim \Gamma(\alpha, 1)$ and $Y \sim \Gamma(\beta, 1)$, then $\frac{X}{Y} \sim \beta'(\alpha, \beta)$.
- $\beta'(p, 1, a, b) = \text{Dagum}(p, a, b)$ the **Dagum distribution**
- $\beta'(1, p, a, b) = \text{SinghMaddala}(p, a, b)$ the **Singh-Maddala distribution**
- $\beta'(1, 1, \gamma, \sigma) = \text{LL}(\gamma, \sigma)$ the **Log logistic distribution**
- Beta prime distribution is a special case of the type 6 **Pearson distribution**
- **Pareto distribution type II** is related to Beta prime distribution
- **Pareto distribution type IV** is related to Beta prime distribution
- **inverted Dirichlet distribution**, a generalization of the beta prime distribution

2.1.4 Notes

[1] Johnson et al (1995), p248

[2] Dubey, Satya D. (December 1970). "Compound gamma, beta and F distributions". *Metrika* **16**: 27–31. doi:10.1007/BF02613934.

2.1.5 References

- Johnson, N.L., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Volume 2 (2nd Edition), Wiley. ISBN 0-471-58494-0
- MathWorld article

2.2 Birnbaum–Saunders distribution

The **Birnbaum–Saunders distribution**, also known as the **fatigue life distribution**, is a probability distribution used extensively in **reliability** applications to model failure times. There are several alternative formulations of this distribution in the literature. It is named after Z. W. Birnbaum and S. C. Saunders.

2.2.1 Theory

This distribution was developed to model failures due to cracks. A material is placed under repeated cycles of stress. The j^{th} cycle leads to an increase in the crack by X_j amount. The sum of the X_j is assumed to be **normally distributed** with mean $n\mu$ and variance $n\sigma^2$. The probability that the crack does not exceed a critical length ω is

$$P(X \leq \omega) = \Phi \left(\frac{\omega - n\mu}{\sigma\sqrt{n}} \right)$$

where $\Phi()$ is the cdf of normal distribution.

If T is the number of cycles to failure then the cumulative distribution function (cdf) of T is

$$P(T \leq t) = 1 - \Phi \left(\frac{\omega - t\mu}{\sigma\sqrt{t}} \right) = \Phi \left(\frac{t\mu - \omega}{\sigma\sqrt{t}} \right) = \Phi \left(\frac{\mu\sqrt{t}}{\sigma} - \frac{\omega}{\sigma\sqrt{t}} \right) = \Phi \left(\frac{\sqrt{\mu\omega}}{\sigma} \left[\left(\frac{t}{\omega/\mu} \right)^{0.5} - \left(\frac{\omega/\mu}{t} \right)^{0.5} \right] \right)$$

The more usual form of this distribution is:

$$F(x; \alpha, \beta) = \Phi \left(\frac{1}{\alpha} \left[\left(\frac{x}{\beta} \right)^{0.5} - \left(\frac{\beta}{x} \right)^{0.5} \right] \right)$$

Here α is the **shape parameter** and β is the **location parameter**.

2.2.2 Properties

The Birnbaum–Saunders distribution is **unimodal** with a **median** of β .

The **mean** (μ), **variance** (σ^2), **skewness** (γ) and **kurtosis** (κ) are as follows:

$$\mu = \beta \left(1 + \frac{\alpha^2}{2} \right)$$

$$\sigma^2 = (\alpha\beta)^2 \left(1 + \frac{5\alpha^2}{4} \right)$$

$$\gamma = \frac{16\alpha^2(11\alpha^2 + 6)}{(5\alpha^2 + 4)^3}$$

$$\kappa = 3 + \frac{6\alpha^2(93\alpha^2 + 41)}{(5\alpha^2 + 4)^2}$$

Given a data set that is thought to be Birnbaum-Saunders distributed the parameters' values are best estimated by **maximum likelihood**.

Differential equation

The cdf of the Birnbaum-Saunders distribution is a solution of the following **differential equation**:

$$\left\{ \begin{array}{l} 2\alpha^2\beta x^2(\beta + x)f'(x) + f(x)(-\beta^3 + x^3 + (\alpha^2 + 1)\beta x^2 + (3\alpha^2 - 1)\beta^2 x) = 0, \\ f(1) = \frac{(\beta+1)e^{-\frac{(\beta-1)^2}{2\alpha^2\beta}}}{2\sqrt{2\pi}\alpha\sqrt{\beta}} \end{array} \right\}$$

If T is Birnbaum-Saunders distributed with parameters α and β then T^{-1} is also Birnbaum-Saunders distributed with parameters α and β^{-1} .

Transformation

Let T be a Birnbaum-Saunders distributed variate with parameters α and β . A useful transformation of T is

$$X = \frac{1}{2} \left[\left(\frac{T}{\beta} \right)^{0.5} - \left(\frac{T}{\beta} \right)^{-0.5} \right]$$

Equivalently

$$T = \beta (1 + 2X^2 + 2X(1 + X^2)^{0.5})$$

X is then distributed normally with a mean of zero and a variance of $\alpha^2 / 4$.

2.2.3 Probability density function

The general formula for the probability density function (pdf) is

$$f(x) = \frac{\sqrt{\frac{x-\mu}{\beta}} + \sqrt{\frac{\beta}{x-\mu}}}{2\gamma(x-\mu)} \phi \left(\frac{\sqrt{\frac{x-\mu}{\beta}} - \sqrt{\frac{\beta}{x-\mu}}}{\gamma} \right) \quad x > \mu; \gamma, \beta > 0$$

where γ is the shape parameter, μ is the location parameter, β is the scale parameter, and ϕ is the probability density function of the standard normal distribution.

2.2.4 Standard fatigue life distribution

The case where $\mu = 0$ and $\beta = 1$ is called the **standard fatigue life distribution**. The pdf for the standard fatigue life distribution reduces to

$$f(x) = \frac{\sqrt{x} + \sqrt{\frac{1}{x}}}{2\gamma x} \phi \left(\frac{\sqrt{x} - \sqrt{\frac{1}{x}}}{\gamma} \right) \quad x > 0; \gamma > 0$$

Since the general form of probability functions can be expressed in terms of the standard distribution, all of the subsequent formulas are given for the standard form of the function.

2.2.5 Cumulative distribution function

The formula for the cumulative distribution function is

$$F(x) = \Phi \left(\frac{\sqrt{x} - \sqrt{\frac{1}{x}}}{\gamma} \right) \quad x > 0; \gamma > 0$$

where Φ is the cumulative distribution function of the standard normal distribution.

2.2.6 Quantile function

The formula for the **quantile function** is

$$G(p) = \frac{1}{4} \left[\gamma \Phi^{-1}(p) + \sqrt{4 + (\gamma \Phi^{-1}(p))^2} \right]^2$$

where Φ^{-1} is the quantile function of the standard normal distribution.

2.2.7 External links

- **Fatigue life distribution**

2.2.8 References

- Birnbaum, Z. W.; Saunders, S. C. (1969), “A new family of life distributions”, *Journal of Applied Probability* **6** (2): 319–327, doi:10.2307/3212003, JSTOR 3212003
- Desmond, A.F. (1985), “Stochastic models of failure in random environments”, *Canadian Journal of Statistics* **13** (3): 171–183, doi:10.2307/3315148, JSTOR 3315148
- Johnson, N.; Kotz, S.; Balakrishnan, N. (1995), *Continuous Univariate Distributions* **2** (2nd ed.), New York: Wiley
- Lemonte, A. J.; Cribari-Neto, F.; Vasconcellos, K. L. P. (2007), “Improved statistical inference for the two-parameter Birnbaum–Saunders distribution”, *Computational Statistics and Data Analysis* **51**: 4656–4681, doi:10.1016/j.csda.2006.08.
- Lemonte, A. J.; Simas, A. B.; Cribari-Neto, F. (2008), “Bootstrap-based improved estimators for the two-parameter Birnbaum–Saunders distribution”, *Journal of Statistical Computation and Simulation* **78**: 37–49, doi:10.1080/10629360600903882
- Cordeiro, G. M.; Lemonte, A. J. (2011), “The β -Birnbaum–Saunders distribution: An improved distribution for fatigue life modeling”, *Computational Statistics and Data Analysis* **55**: 1445–1461, doi:10.1016/j.csda.2010.10.007
- Lemonte, A. J. (2013), “A new extension of the Birnbaum–Saunders distribution”, *Brazilian Journal of Probability and Statistics* **27**: 133–149, doi:10.1214/11-BJPS160

This article incorporates **public domain material** from websites or documents of the **National Institute of Standards and Technology**.

2.3 Chi distribution

See also: **Chi-squared distribution**

In **probability theory** and **statistics**, the **chi distribution** is a continuous **probability distribution**. It is the distribution of the square root of the sum of squares of independent random variables having a **standard normal distribution**. The most familiar examples are the **Rayleigh distribution** with chi distribution with 2 degrees of freedom, and the **Maxwell distribution** of (normalized) molecular speeds which is a chi distribution with 3 degrees of freedom (one for each spatial coordinate). If X_i are k independent, **normally distributed** random variables with means μ_i and **standard deviations** σ_i , then the statistic

$$Y = \sqrt{\sum_{i=1}^k \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2}$$

is distributed according to the chi distribution. Accordingly, dividing by the mean of the chi distribution (scaled by the square root of $n - 1$) yields the correction factor in the unbiased estimation of the standard deviation of the normal distribution. The chi distribution has one parameter: k which specifies the number of degrees of freedom (i.e. the number of X_i).

2.3.1 Characterization

Probability density function

The probability density function is

$$f(x; k) = \frac{2^{1-\frac{k}{2}} x^{k-1} e^{-\frac{x^2}{2}}}{\Gamma(\frac{k}{2})}$$

where $\Gamma(z)$ is the Gamma function.

Cumulative distribution function

The cumulative distribution function is given by:

$$F(x; k) = P(k/2, x^2/2)$$

where $P(k, x)$ is the regularized Gamma function.

Generating functions

Moment generating function The moment generating function is given by:

$$M(t) = M\left(\frac{k}{2}, \frac{1}{2}, \frac{t^2}{2}\right) + t\sqrt{2} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} M\left(\frac{k+1}{2}, \frac{3}{2}, \frac{t^2}{2}\right)$$

Characteristic function The characteristic function is given by:

$$\varphi(t; k) = M\left(\frac{k}{2}, \frac{1}{2}, \frac{-t^2}{2}\right) + it\sqrt{2} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} M\left(\frac{k+1}{2}, \frac{3}{2}, \frac{-t^2}{2}\right)$$

where again, $M(a, b, z)$ is Kummer's confluent hypergeometric function.

2.3.2 Properties

Differential equation

$$\left\{ x f'(x) + f(x) (-\nu + x^2 + 1) = 0, f(1) = \frac{2^{1-\frac{\nu}{2}}}{\sqrt{e}\Gamma(\frac{\nu}{2})} \right\}$$

Moments

The raw **moments** are then given by:

$$\mu_j = 2^{j/2} \frac{\Gamma((k+j)/2)}{\Gamma(k/2)}$$

where $\Gamma(z)$ is the **Gamma function**. The first few raw moments are:

$$\mu_1 = \sqrt{2} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)}$$

$$\mu_2 = k$$

$$\mu_3 = 2\sqrt{2} \frac{\Gamma((k+3)/2)}{\Gamma(k/2)} = (k+1)\mu_1$$

$$\mu_4 = (k)(k+2)$$

$$\mu_5 = 4\sqrt{2} \frac{\Gamma((k+5)/2)}{\Gamma(k/2)} = (k+1)(k+3)\mu_1$$

$$\mu_6 = (k)(k+2)(k+4)$$

where the rightmost expressions are derived using the recurrence relationship for the Gamma function:

$$\Gamma(x+1) = x\Gamma(x)$$

From these expressions we may derive the following relationships:

$$\text{Mean: } \mu = \sqrt{2} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)}$$

$$\text{Variance: } \sigma^2 = k - \mu^2$$

$$\text{Skewness: } \gamma_1 = \frac{\mu}{\sigma^3} (1 - 2\sigma^2)$$

$$\text{Kurtosis excess: } \gamma_2 = \frac{2}{\sigma^2} (1 - \mu\sigma\gamma_1 - \sigma^2)$$

Entropy

The entropy is given by:

$$S = \ln(\Gamma(k/2)) + \frac{1}{2}(k - \ln(2) - (k-1)\psi_0(k/2))$$

where $\psi_0(z)$ is the **polygamma function**.

2.3.3 Related distributions

- If $X \sim \chi_k(x)$ then $X^2 \sim \chi_k^2$ (**chi-squared distribution**)
- $\lim_{k \rightarrow \infty} \frac{\chi_k(x) - \mu_k}{\sigma_k} \xrightarrow{d} N(0, 1)$ (**Normal distribution**)
- If $X \sim N(0, 1)$ then $|X| \sim \chi_1(x)$

- If $X \sim \chi_1(x)$ then $\sigma X \sim HN(\sigma)$ (half-normal distribution) for any $\sigma > 0$
- $\chi_2(x) \sim \text{Rayleigh}(1)$ (Rayleigh distribution)
- $\chi_3(x) \sim \text{Maxwell}(1)$ (Maxwell distribution)
- $\|N_{i=1, \dots, k}(0, 1)\|_2 \sim \chi_k(x)$ (The 2-norm of k standard normally distributed variables is a chi distribution with k degrees of freedom)
- chi distribution is a special case of the generalized gamma distribution or the nakagami distribution or the noncentral chi distribution

2.3.4 See also

- Nakagami distribution

2.3.5 External links

- <http://mathworld.wolfram.com/ChiDistribution.html>

2.4 Chi-squared distribution

This article is about the mathematics of the chi-squared distribution. For its uses in statistics, see **chi-squared test**. For the music group, see **Chi2 (band)**.

In probability theory and statistics, the **chi-squared distribution** (also **chi-square** or **χ^2 -distribution**) with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. It is a special case of the gamma distribution and is one of the most widely used probability distributions in inferential statistics, e.g., in hypothesis testing or in construction of confidence intervals.^{[2][3][4][5]} When it is being distinguished from the more general noncentral chi-squared distribution, this distribution is sometimes called the **central chi-squared distribution**.

The chi-squared distribution is used in the common chi-squared tests for goodness of fit of an observed distribution to a theoretical one, the independence of two criteria of classification of qualitative data, and in confidence interval estimation for a population standard deviation of a normal distribution from a sample standard deviation. Many other statistical tests also use this distribution, like Friedman's analysis of variance by ranks.

2.4.1 Definition

If $Z_1, \dots, Z_k$ are independent, standard normal random variables, then the sum of their squares,

$$Q = \sum_{i=1}^k Z_i^2,$$

is distributed according to the **chi-squared distribution** with k degrees of freedom. This is usually denoted as

$$Q \sim \chi^2(k) \text{ or } Q \sim \chi_k^2.$$

The chi-squared distribution has one parameter: k — a positive integer that specifies the number of degrees of freedom (i.e. the number of Z_i 's)

2.4.2 Introduction to the chi-squared distribution

The chi-squared distribution is used primarily in hypothesis testing. Unlike more widely-known distributions such as the **normal distribution** and the **exponential distribution**, the chi-squared distribution is rarely used to model natural phenomena. It arises in the following hypothesis tests, among others.

- **Chi-squared test** of independence in contingency tables
- **Chi-squared test** of goodness of fit of observed data to hypothetical distributions
- **Likelihood-ratio test** for nested models
- **Log-rank test** in survival analysis
- **Cochran–Mantel–Haenszel test** for stratified contingency tables

It is also a component of the definition of the **t-distribution** and the **F-distribution** used in t-tests, analysis of variance, and regression analysis.

The primary reason that the chi-squared distribution is used extensively in hypothesis testing is its relationship to the normal distribution. Many hypothesis tests use a test statistic, such as the **t statistic** in a t-test. For these hypothesis tests, as the sample size, n , increases, the **sampling distribution** of the test statistic approaches the normal distribution (**Central Limit Theorem**). Because the test statistic (such as t) is asymptotically normally distributed, provided the sample size is sufficiently large, the distribution used for hypothesis testing may be approximated by a normal distribution. Testing hypotheses using a normal distribution is well understood and relatively easy. The simplest chi-squared distribution is the square of a standard normal distribution. So wherever a normal distribution could be used for a hypothesis test, a chi-squared distribution could be used.

Specifically, suppose that Z is a standard normal random variable, with mean = 0 and variance = 1. $Z \sim N(0,1)$. A sample drawn at random from Z is a sample from the distribution shown in the graph of the standard **normal distribution**. Define a new random variable Q . To generate a random sample from Q , take a sample from Z and square the value. The distribution of the squared values is given by the random variable $Q = Z^2$. The distribution of the random variable Q is an example of a chi-squared distribution: $Q \sim \chi_1^2$. The subscript 1 indicates that this particular chi-squared distribution is constructed from only 1 standard normal distribution. A chi-squared distribution constructed by squaring a single standard normal distribution is said to have 1 degree of freedom. Thus, as the sample size for a hypothesis test increases, the distribution of the test statistic approaches a normal distribution. and the distribution of the *square* of the test statistic approaches a chi-squared distribution. Just as extreme values of the normal distribution have low probability (and give small p-values), extreme values of the chi-squared distribution have low probability.

An additional reason that the chi-squared distribution is widely used is that it is a member of the class of **likelihood ratio tests (LRT)**.^[6] LRT's have several desirable properties; in particular, LRT's commonly provide the highest power to reject the null hypothesis (**Neyman–Pearson lemma**). However, the normal and chi-squared approximations are only valid asymptotically. For this reason, it is preferable to use the t distribution rather than the normal approximation or the chi-squared approximation for small sample size. Similarly, in analyses of contingency tables, the chi-squared approximation will be poor for small sample size, and it is preferable to use the Fisher Exact test. Ramsey and Ramsey show that the exact binomial test is always more powerful than the normal approximation.^[7]

Where are the squared normal distributions?

If the chi-squared distribution is used because it is the sum of squared normal distributions, where are the squared normal distributions in contingency tables analyzed with a chi-squared test? The answer can be traced back to the normal approximation to the binomial distribution. Consider an experiment in which 10 fair coins are tossed, and the number of heads is observed. This experiment can be modeled with a binomial distribution, with $n=10$ trials and $p = 0.5$ probability of heads on each trial. Suppose that heads is observed 1 times in 10 trials. What is the probability of a result as extreme as 1 heads in 10 trials, if the probability of heads is $p=0.5$?

Three methods to determine the probability are:

- Calculate the probability exactly using the binomial distribution.
- Estimate the probability using normal approximation to the binomial distribution.
- Estimate the probability using a chi-squared test. This result will be the same as the result for the normal approximation.

Calculation using the exact binomial and the normal approximation may be performed using <http://vassarstats.net/binomialX.html>. Calculation of the chi-square probability may be performed using <http://vassarstats.net/csfit.html>.

Using the binomial distribution, the probability of a result as extreme 1 heads in 10 trials is the sum of the probabilities of 0 heads, 1 head, 9 heads, or 10 heads. Notice that this is a two-tailed or two-sided test. This test gives $p=0.0215$.

Using the normal approximation to the binomial distribution, the (two-sided) probability of a result as extreme 1 heads in 10 trials is $p=0.0271$.

The chi-squared test is performed as follows. The observed number of heads is 1, and the observed number of tails is 9. The expected number of heads = expected number of tails = $10 \cdot 0.5 = 5$. The difference between the observed and expected is $1-5=-4$ for heads, and $9-5=4$ for tails. The chi-squared statistic (with Yates's correction for continuity) is

$$\chi^2 = \frac{(|1 - 5| - 0.5)^2}{5} + \frac{(|9 - 5| - 0.5)^2}{5} = 4.9.$$

For the chi-squared test, the (two-sided) probability of a result as extreme as 1 heads in 10 trials is $p=0.027$, the same as the result using the normal approximation. That is, the probability that the chi-squared statistic with one degree of freedom is greater than 4.9 is $p=0.027$.

Lancaster^[8] shows the connections among the binomial, normal, and chi-squared distributions, as follows. De Moivre and Laplace established that a binomial distribution could be approximated by a normal distribution. Specifically they showed the asymptotic normality of the random variable

$$\chi = \frac{(m - Np)}{\sqrt{Npq}}$$

where m is the observed number of successes in N trials, where the probability of success is p , and $q = 1 - p$.

Squaring both sides of the equation gives

$$\chi^2 = \frac{(m - Np)^2}{(Npq)}$$

Using $N = Np + N(1 - p)$, $N = m + (N - m)$, and $q = 1 - p$, this equation simplifies to

$$\chi^2 = \frac{(m - Np)^2}{(Np)} + \frac{(N - m - Nq)^2}{(Nq)}$$

The expression on the right is of the form that Pearson would generalize to the form:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

where

χ^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution.

O_i = the number of observations of type i .

$E_i = Np_i$ = the expected (theoretical) frequency of type i , asserted by the null hypothesis that the fraction of type i in the population is p_i

n = the number of cells in the table.

In the case of a binomial outcome (flipping a coin), the binomial distribution may be approximated by a normal distribution (for sufficiently large n). Because the square of a standard normal distribution is the chi-squared distribution with one degree of freedom, the probability of a result such as 1 heads in 10 trials can be approximated either by the normal or the chi-squared distribution. However, many problems involve more than the two possible outcomes of a binomial, and instead require 3 or more categories, which leads to the multinomial distribution. Just as de Moivre and Laplace sought for and found the normal approximation to the binomial, Pearson sought for and found a multivariate normal approximation to the multinomial distribution. Pearson showed that the chi-squared distribution, the sum of multiple normal distributions, was such an approximation to the multinomial distribution ^[8]

2.4.3 Characteristics

Further properties of the chi-squared distribution can be found in the box at the upper right corner of this article.

Probability density function

The **probability density function** (pdf) of the chi-squared distribution is

$$f(x; k) = \begin{cases} \frac{x^{(k/2-1)} e^{-x/2}}{2^{k/2} \Gamma(\frac{k}{2})}, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

where $\Gamma(k/2)$ denotes the **Gamma function**, which has closed-form values for integer k .

For derivations of the pdf in the cases of one, two and k degrees of freedom, see **Proofs related to chi-squared distribution**.

Differential equation

The pdf of the chi-squared distribution is a solution to the following **differential equation**:

$$\left\{ \begin{array}{l} 2xf'(x) + f(x)(-k + x + 2) = 0, \\ f(1) = \frac{2^{-k/2}}{\sqrt{e}\Gamma(\frac{k}{2})} \end{array} \right\}$$

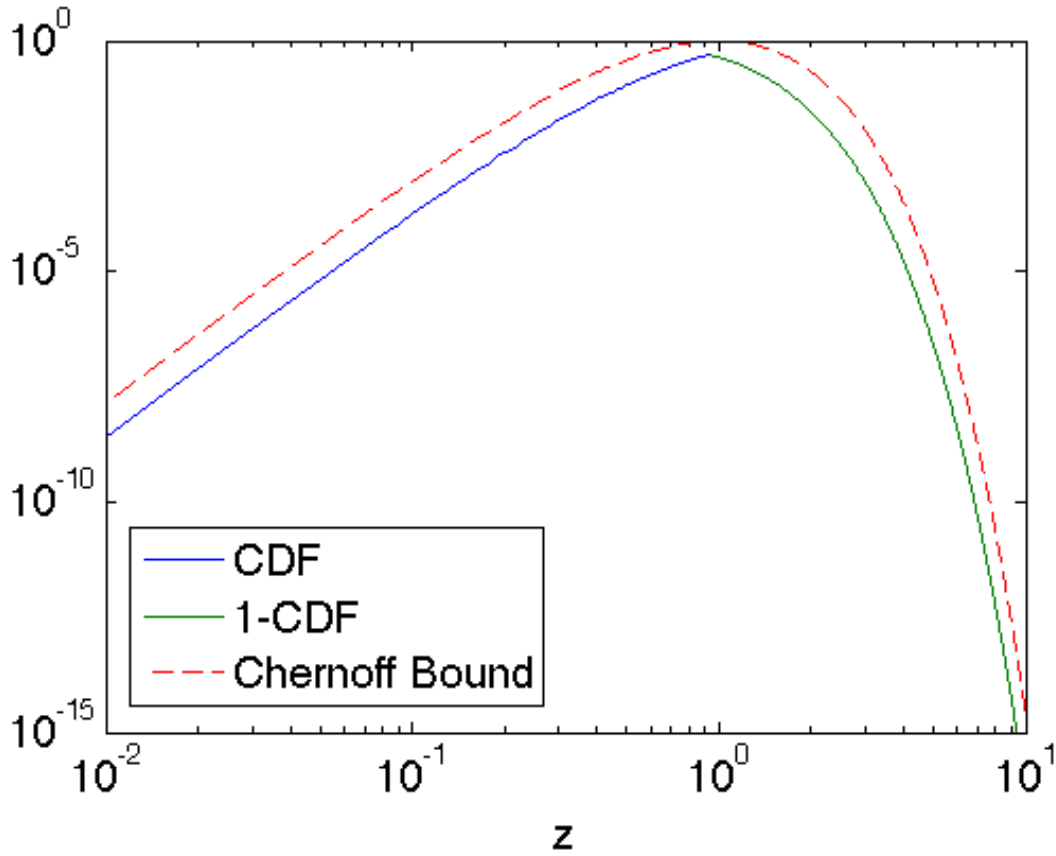
Cumulative distribution function

Its **cumulative distribution function** is:

$$F(x; k) = \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(\frac{k}{2})} = P\left(\frac{k}{2}, \frac{x}{2}\right),$$

where $\gamma(s, t)$ is the **lower incomplete Gamma function** and $P(s, t)$ is the **regularized Gamma function**.

In a special case of $k = 2$ this function has a simple form:



Chernoff bound for the CDF and tail (1-CDF) of a chi-squared random variable with ten degrees of freedom ($k = 10$)

$$F(x; 2) = 1 - e^{-\frac{x}{2}}$$

and the form is not much more complicated for other small even k .

Tables of the chi-squared cumulative distribution function are widely available and the function is included in many spreadsheets and all statistical packages.

Letting $z \equiv x/k$, **Chernoff bounds** on the lower and upper tails of the CDF may be obtained.^[9] For the cases when $0 < z < 1$ (which include all of the cases when this CDF is less than half):

$$F(zk; k) \leq (ze^{1-z})^{k/2}.$$

The tail bound for the cases when $z > 1$, similarly, is

$$1 - F(zk; k) \leq (ze^{1-z})^{k/2}.$$

For another approximation for the CDF modeled after the cube of a Gaussian, see under Noncentral chi-squared distribution.

Additivity

It follows from the definition of the chi-squared distribution that the sum of independent chi-squared variables is also chi-squared distributed. Specifically, if $\{X_i\}_{i=1}^n$ are independent chi-squared variables with $\{k_i\}_{i=1}^n$ degrees of freedom, respectively, then $Y = X_1 + \dots + X_n$ is chi-squared distributed with $k_1 + \dots + k_n$ degrees of freedom.

Sample mean

The sample mean of n i.i.d. chi-squared variables of degree k is distributed according to a gamma distribution with shape α and scale θ parameters:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim \text{Gamma}(\alpha = nk/2, \theta = 2/n) \quad \text{where} \quad X_i \sim \chi^2(k)$$

Asymptotically, given that for a scale parameter α going to infinity, a Gamma distribution converges towards a Normal distribution with expectation $\mu = \alpha \cdot \theta$ and variance $\sigma^2 = \alpha \theta^2$, the sample mean converges towards:

$$\bar{X} \xrightarrow{n \rightarrow \infty} N(\mu = k, \sigma^2 = 2k/n)$$

Note that we would have obtained the same result invoking instead the **central limit theorem**, noting that for each chi-squared variable of degree k the expectation is k , and its variance $2k$ (and hence the variance of the sample mean $\bar{X}$ being $\sigma^2 = 2k/n$).

Entropy

The **differential entropy** is given by

$$h = \int_{-\infty}^{\infty} f(x; k) \ln f(x; k) dx = \frac{k}{2} + \ln \left[2 \Gamma \left(\frac{k}{2} \right) \right] + \left(1 - \frac{k}{2} \right) \psi \left[\frac{k}{2} \right],$$

where $\psi(x)$ is the Digamma function.

The chi-squared distribution is the **maximum entropy probability distribution** for a random variate X for which $E(X) = k$ and $E(\ln(X)) = \psi(k/2) + \log(2)$ are fixed. Since the chi-squared is in the family of gamma distributions, this can be derived by substituting appropriate values in the **Expectation of the Log moment of Gamma**. For derivation from more basic principles, see the derivation in **moment generating function of the sufficient statistic**.

Noncentral moments

The moments about zero of a chi-squared distribution with k degrees of freedom are given by^{[10][11]}

$$E(X^m) = k(k+2)(k+4) \cdots (k+2m-2) = 2^m \frac{\Gamma(m + \frac{k}{2})}{\Gamma(\frac{k}{2})}.$$

Cumulants

The **cumulants** are readily obtained by a (formal) power series expansion of the logarithm of the characteristic function:

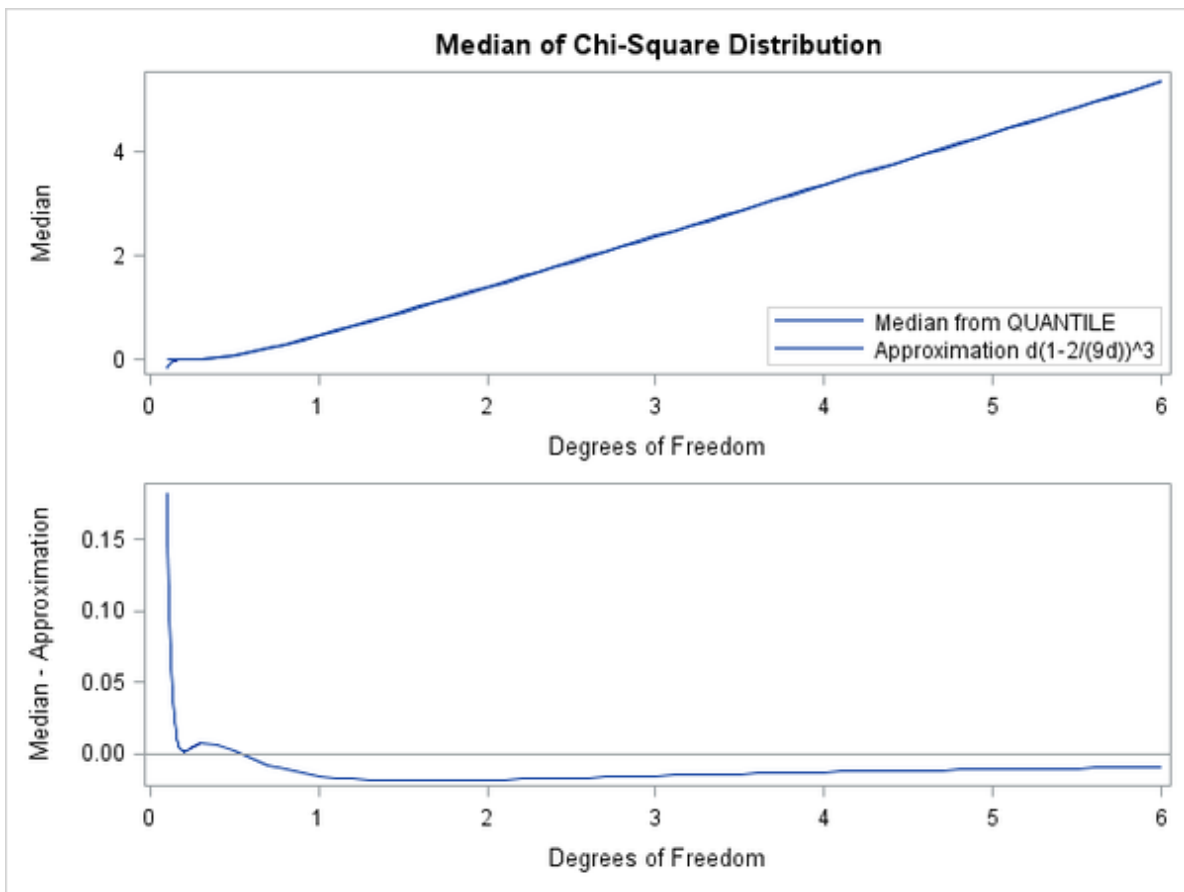
$$\kappa_n = 2^{n-1} (n-1)! k$$

Asymptotic properties

By the **central limit theorem**, because the chi-squared distribution is the sum of k independent random variables with finite mean and variance, it converges to a normal distribution for large k . For many practical purposes, for $k > 50$ the distribution is sufficiently close to a **normal distribution** for the difference to be ignored.^[12] Specifically, if $X \sim \chi^2(k)$, then as k tends to infinity, the distribution of $(X - k)/\sqrt{2k}$ tends to a standard normal distribution. However, convergence is slow as the **skewness** is $\sqrt{8/k}$ and the **excess kurtosis** is $12/k$.

- The sampling distribution of $\ln(\chi^2)$ converges to normality much faster than the sampling distribution of χ^2 ,^[13] as the logarithm removes much of the asymmetry.^[14] Other functions of the chi-squared distribution converge more rapidly to a normal distribution. Some examples are:
- If $X \sim \chi^2(k)$ then $\sqrt{2X}$ is approximately normally distributed with mean $\sqrt{2k-1}$ and unit variance (result credited to **R. A. Fisher**).
- If $X \sim \chi^2(k)$ then $\sqrt[3]{X/k}$ is approximately normally distributed with mean $1-2/(9k)$ and variance $2/(9k)$.^[15] This is known as the Wilson–Hilferty transformation.

2.4.4 Relation to other distributions



Approximate formula for median compared with numerical quantile (top). Difference between numerical quantile and approximate formula (bottom).

- As $k \rightarrow \infty$, $(\chi_k^2 - k)/\sqrt{2k} \xrightarrow{d} N(0, 1)$ (normal distribution)

- $\chi_k^2 \sim \chi_k'^2(0)$ (**Noncentral chi-squared distribution** with non-centrality parameter $\lambda = 0$)
- If $X \sim F(\nu_1, \nu_2)$ then $Y = \lim_{\nu_2 \rightarrow \infty} \nu_1 X$ has the chi-squared distribution $\chi_{\nu_1}^2$
- As a special case, if $X \sim F(1, \nu_2)$ then $Y = \lim_{\nu_2 \rightarrow \infty} X$ has the chi-squared distribution χ_1^2
- $\|N_{i=1, \dots, k}(0, 1)\|^2 \sim \chi_k^2$ (The squared **norm** of **k** standard normally distributed variables is a chi-squared distribution with **k** **degrees of freedom**)
- If $X \sim \chi^2(\nu)$ and $c > 0$, then $cX \sim \Gamma(k = \nu/2, \theta = 2c)$. (**gamma distribution**)
- If $X \sim \chi_k^2$ then $\sqrt{X} \sim \chi_k$ (**chi distribution**)
- If $X \sim \chi^2(2)$, then $X \sim \text{Exp}(1/2)$ is an **exponential distribution**. (See **Gamma distribution** for more.)
- If $X \sim \text{Rayleigh}(1)$ (**Rayleigh distribution**) then $X^2 \sim \chi^2(2)$
- If $X \sim \text{Maxwell}(1)$ (**Maxwell distribution**) then $X^2 \sim \chi^2(3)$
- If $X \sim \chi^2(\nu)$ then $\frac{1}{X} \sim \text{Inv-}\chi^2(\nu)$ (**Inverse-chi-squared distribution**)
- The chi-squared distribution is a special case of type 3 **Pearson distribution**
- If $X \sim \chi^2(\nu_1)$ and $Y \sim \chi^2(\nu_2)$ are independent then $\frac{X}{X+Y} \sim \text{Beta}(\frac{\nu_1}{2}, \frac{\nu_2}{2})$ (**beta distribution**)
- If $X \sim U(0, 1)$ (**uniform distribution**) then $-2 \log(X) \sim \chi^2(2)$
- $\chi^2(6)$ is a transformation of **Laplace distribution**
- If $X_i \sim \text{Laplace}(\mu, \beta)$ then $\sum_{i=1}^n \frac{2|X_i - \mu|}{\beta} \sim \chi^2(2n)$
- chi-squared distribution is a transformation of **Pareto distribution**
- **Student's t-distribution** is a transformation of chi-squared distribution
- **Student's t-distribution** can be obtained from chi-squared distribution and **normal distribution**
- **Noncentral beta distribution** can be obtained as a transformation of chi-squared distribution and **Noncentral chi-squared distribution**
- **Noncentral t-distribution** can be obtained from normal distribution and chi-squared distribution

A chi-squared variable with k degrees of freedom is defined as the sum of the squares of k independent **standard normal** random variables.

If Y is a k -dimensional Gaussian random vector with mean vector μ and rank k covariance matrix C , then $X = (Y - \mu)^T C^{-1} (Y - \mu)$ is chi-squared distributed with k degrees of freedom.

The sum of squares of **statistically independent** unit-variance Gaussian variables which do *not* have mean zero yields a generalization of the chi-squared distribution called the **noncentral chi-squared distribution**.

If Y is a vector of k i.i.d. standard normal random variables and A is a $k \times k$ **symmetric, idempotent matrix** with rank $k - n$ then the **quadratic form** $Y^T A Y$ is chi-squared distributed with $k - n$ degrees of freedom.

The chi-squared distribution is also naturally related to other distributions arising from the Gaussian. In particular,

- Y is **F-distributed**, $Y \sim F(k_1, k_2)$ if $Y = \frac{X_1/k_1}{X_2/k_2}$ where $X_1 \sim \chi^2(k_1)$ and $X_2 \sim \chi^2(k_2)$ are statistically independent.
- If X is chi-squared distributed, then $\sqrt{X}$ is **chi distributed**.
- If $X_1 \sim \chi^2 k_1$ and $X_2 \sim \chi^2 k_2$ are statistically independent, then $X_1 + X_2 \sim \chi^2 k_1 + k_2$. If X_1 and X_2 are not independent, then $X_1 + X_2$ is not chi-squared distributed.

2.4.5 Generalizations

The chi-squared distribution is obtained as the sum of the squares of k independent, zero-mean, unit-variance Gaussian random variables. Generalizations of this distribution can be obtained by summing the squares of other types of Gaussian random variables. Several such distributions are described below.

Linear combination

If $X_1, \dots, X_n$ are chi square random variables and $a_1, \dots, a_n \in \mathbb{R}_{>0}$, then a closed expression for the distribution of $X = \sum_{i=1}^n a_i X_i$ is not known. It may be, however, calculated using the [property of characteristic functions](#) of the chi-squared random variable.^[16]

Chi-squared distributions

Noncentral chi-squared distribution Main article: [Noncentral chi-squared distribution](#)

The noncentral chi-squared distribution is obtained from the sum of the squares of independent Gaussian random variables having unit variance and *nonzero* means.

Generalized chi-squared distribution Main article: [Generalized chi-squared distribution](#)

The generalized chi-squared distribution is obtained from the quadratic form $z'Az$ where z is a zero-mean Gaussian vector having an arbitrary covariance matrix, and A is an arbitrary matrix.

Gamma, exponential, and related distributions

The chi-squared distribution $X \sim \chi^2(k)$ is a special case of the [gamma distribution](#), in that $X \sim \Gamma(k/2, 1/2)$ using the rate parameterization of the gamma distribution (or $X \sim \Gamma(k/2, 2)$ using the scale parameterization of the gamma distribution) where k is an integer.

Because the [exponential distribution](#) is also a special case of the Gamma distribution, we also have that if $X \sim \chi^2(2)$, then $X \sim \text{Exp}(1/2)$ is an [exponential distribution](#).

The [Erlang distribution](#) is also a special case of the Gamma distribution and thus we also have that if $X \sim \chi^2(k)$ with even k , then X is Erlang distributed with shape parameter $k/2$ and scale parameter $1/2$.

2.4.6 Applications

The chi-squared distribution has numerous applications in inferential [statistics](#), for instance in [chi-squared tests](#) and in estimating [variances](#). It enters the problem of estimating the mean of a normally distributed population and the problem of estimating the slope of a [regression](#) line via its role in [Student's t-distribution](#). It enters all analysis of [variance](#) problems via its role in the [F-distribution](#), which is the distribution of the ratio of two independent chi-squared [random variables](#), each divided by their respective degrees of freedom.

Following are some of the most common situations in which the chi-squared distribution arises from a Gaussian-distributed sample.

- if $X_1, \dots, X_n$ are i.i.d. $N(\mu, \sigma^2)$ [random variables](#), then $\sum_{i=1}^n (X_i - \bar{X})^2 \sim \sigma^2 \chi_{n-1}^2$ where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.
- The box below shows some [statistics](#) based on $X_i \sim \text{Normal}(\mu_i, \sigma^2 i)$, $i = 1, \dots, k$, independent random variables that have probability distributions related to the chi-squared distribution:

The chi-squared distribution is also often encountered in Magnetic Resonance Imaging .^[17]

2.4.7 Table of χ^2 value vs p -value

The p -value is the probability of observing a test statistic *at least* as extreme in a chi-squared distribution. Accordingly, since the cumulative distribution function (CDF) for the appropriate degrees of freedom (df) gives the probability of having obtained a value *less extreme* than this point, subtracting the CDF value from 1 gives the p -value. The table below gives a number of p -values matching to χ^2 for the first 10 degrees of freedom.

A low p -value indicates greater statistical significance, i.e. greater confidence that the observed deviation from the null hypothesis is significant. A p -value of 0.05 is often used as a cutoff between significant and not-significant results.

2.4.8 History and name

This distribution was first described by the German statistician Friedrich Robert Helmert in papers of 1875-6,^{[19][20]} where he computed the sampling distribution of the sample variance of a normal population. Thus in German this was traditionally known as the *Helmert'sche* (“Helmertian”) or “Helmert distribution”.

The distribution was independently rediscovered by the English mathematician Karl Pearson in the context of goodness of fit, for which he developed his Pearson's chi-squared test, published in 1900, with computed table of values published in (Elderton 1902), collected in (Pearson 1914, pp. xxxi–xxxiii, 26–28, Table XII). The name “chi-squared” ultimately derives from Pearson's shorthand for the exponent in a multivariate normal distribution with the Greek letter Chi, writing $-\frac{1}{2}\chi^2$ for what would appear in modern notation as $-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}$ (Σ being the covariance matrix).^[21] The idea of a family of “chi-squared distributions”, however, is not due to Pearson but arose as a further development due to Fisher in the 1920s.^[19]

2.4.9 See also

- Cochran's theorem
- F -distribution
- Fisher's method for combining independent tests of significance
- Gamma distribution
- Generalized chi-squared distribution
- Hotelling's T -squared distribution
- Noncentral chi-squared distribution
- Pearson's chi-squared test
- Student's t -distribution
- Wilks' lambda distribution
- Wishart distribution

2.4.10 References

- [1] M.A. Sanders. “Characteristic function of the central chi-squared distribution” (PDF). Retrieved 2009-03-06.
- [2] Abramowitz, Milton; Stegun, Irene A., eds. (December 1972) [1964]. “Chapter 26”. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Applied Mathematics Series **55** (10 ed.). New York, USA: United States Department of Commerce, National Bureau of Standards; Dover Publications. p. 940. ISBN 978-0-486-61272-0. LCCN 64-60036. MR 0167642.
- [3] NIST (2006). *Engineering Statistics Handbook - Chi-Squared Distribution*
- [4] Jonhson, N. L.; Kotz, S.; Balakrishnan, N. (1994). “Chi-Squared Distributions including Chi and Rayleigh”. *Continuous Univariate Distributions* **1** (Second ed.). John Wiley and Sons. pp. 415–493. ISBN 0-471-58495-9.
- [5] Mood, Alexander; Graybill, Franklin A.; Boes, Duane C. (1974). *Introduction to the Theory of Statistics* (Third ed.). McGraw-Hill. pp. 241–246. ISBN 0-07-042864-6.
- [6] Westfall, Peter H. (2013). *Understanding Advanced Statistical Methods*. Boca Raton, FL: CRC Press. ISBN 978-1-4665-1210-8.
- [7] Ramsey, PH (1988). “Evaluating the Normal Approximation to the Binomial Test”. *Journal of Educational Statistics* **13** (2): 173–82.
- [8] Lancaster, H.O. (1969), *The Chi-squared Distribution*, Wiley
- [9] Dasgupta, Sanjoy D. A.; Gupta, Anupam K. (2002). “An Elementary Proof of a Theorem of Johnson and Lindenstrauss” (PDF). *Random Structures and Algorithms* **22**: 60–65. doi:10.1002/rsa.10073. Retrieved 2012-05-01.
- [10] Chi-squared distribution, from MathWorld, retrieved Feb. 11, 2009
- [11] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables*, New York: Springer, 2002, eq. (2.35), ISBN 978-0-387-34657-1
- [12] Box, Hunter and Hunter (1978). *Statistics for experimenters*. Wiley. p. 118. ISBN 0471093157.
- [13] Bartlett, M. S.; Kendall, D. G. (1946). “The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation”. *Supplement to the Journal of the Royal Statistical Society* **8** (1): 128–138. JSTOR 2983618.
- [14] Shoemaker, Lewis H. (2003). “Fixing the F Test for Equal Variances”. *The American Statistician* **57** (2): 105–114. doi:10.1198/0003130031441. JSTOR 30037243.
- [15] Wilson, E. B.; Hilferty, M. M. (1931). “The distribution of chi-squared” (PDF). *Proc. Natl. Acad. Sci. USA* **17** (12): 684–688.
- [16] Davies, R.B. (1980). “Algorithm AS155: The Distributions of a Linear Combination of χ^2 Random Variables”. *Journal of the Royal Statistical Society* **29** (3): 323–333. doi:10.2307/2346911.
- [17] den Dekker A. J., Sijbers J., (2014) “Data distributions in magnetic resonance images: a review”, *Physica Medica*,
- [18] Chi-Squared Test Table B.2. Dr. Jacqueline S. McLaughlin at The Pennsylvania State University. In turn citing: R.A. Fisher and F. Yates, *Statistical Tables for Biological Agricultural and Medical Research*, 6th ed., Table IV
- [19] Hald 1998, pp. 633–692, 27. Sampling Distributions under Normality.
- [20] F. R. Helmert, "Ueber die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit im Zusammenhang stehende Fragen", *Zeitschrift für Mathematik und Physik* 21, 1876, S. 102–219
- [21] R. L. Plackett, *Karl Pearson and the Chi-Squared Test*, International Statistical Review, 1983, 61f. See also Jeff Miller, Earliest Known Uses of Some of the Words of Mathematics.

2.4.11 Further reading

- Hald, Anders (1998). *A history of mathematical statistics from 1750 to 1930*. New York: Wiley. ISBN 0-471-17912-4.
- Elderton, William Palin (1902). “Tables for Testing the Goodness of Fit of Theory to Observation”. *Biometrika* **1** (2): 155–163. doi:10.1093/biomet/1.2.155.

2.4.12 External links

- Hazewinkel, Michiel, ed. (2001), “Chi-squared distribution”, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Calculator for the pdf, cdf and quantiles of the chi-squared distribution
- Earliest Uses of Some of the Words of Mathematics: entry on Chi squared has a brief history
- Course notes on Chi-Squared Goodness of Fit Testing from Yale University Stats 101 class.
- *Mathematica* demonstration showing the chi-squared sampling distribution of various statistics, e.g. $\sum x^2$, for a normal population
- Simple algorithm for approximating cdf and inverse cdf for the chi-squared distribution with a pocket calculator

2.5 Dagum distribution

The **Dagum distribution** is a continuous probability distribution defined over positive real numbers. It is named after Camilo Dagum, who proposed it in a series of papers in the 1970s.^{[1][2]} The Dagum distribution arose from several variants of a new model on the size distribution of personal income and is mostly associated with the study of income distribution. There is both a three-parameter specification (Type I) and a four-parameter specification (Type II) of the Dagum distribution; a summary of the genesis of this distribution can be found in “A Guide to the Dagum Distributions”.^[3] A general source on statistical size distributions often cited in work using the Dagum distribution is *Statistical Size Distributions in Economics and Actuarial Sciences*.^[4]

2.5.1 Definition

The cumulative distribution function of the Dagum distribution (Type I) is given by

$$F(x; a, b, p) = \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-p} \text{ for } x > 0 \text{ and where } a, b, p > 0.$$

The corresponding probability density function is given by

$$f(x; a, b, p) = \frac{ap}{x} \left(\frac{\left(\frac{x}{b}\right)^{ap}}{\left(\left(\frac{x}{b}\right)^a + 1\right)^{p+1}} \right).$$

The Dagum distribution can be derived as a special case of the Generalized Beta II (GB2) distribution (a generalization of the Beta prime distribution). There is also an intimate relationship between the Dagum and Singh-Maddala distribution.

$$X \sim D(a, b, p) \iff \frac{1}{X} \sim SM(a, \frac{1}{b}, p)$$

The cumulative distribution function of the Dagum (Type II) distribution adds a point mass at the origin and then follows a Dagum (Type I) distribution over the rest of the support (i.e. over the positive halfline)

$$F(x; a, b, p, \delta) = \delta + (1 - \delta) \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-p}.$$

2.5.2 References

- [1] Dagum, Camilo (1975); A model of income distribution and the conditions of existence of moments of finite order; *Bulletin of the International Statistical Institute*, 46 (Proceeding of the 40th Session of the ISI, Contributed Paper), 199-205.
- [2] Dagum, Camilo (1977); A new model of personal income distribution: Specification and estimation; *Economie Appliquée*, 30, 413-437.
- [3] Kleiber, Christian (2008) “A Guide to the Dagum Distributions” in Chotikapanich, Duangkamon (ed.) *Modeling Income Distributions and Lorenz Curves (Economic Studies in Inequality, Social Exclusion and Well-Being)*, Chapter 6, Springer
- [4] Kleiber, Christian and Samuel Kotz (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley

2.5.3 External links

- Camilo Dagum (1925 - 2005) : obituary

2.6 Exponential distribution

Not to be confused with the **exponential family** of probability distributions.

In probability theory and statistics, the **exponential distribution** (a.k.a. **negative exponential distribution**) is the probability distribution that describes the time between events in a **Poisson process**, i.e. a process in which events occur continuously and independently at a constant average rate. It is a particular case of the **gamma distribution**. It is the continuous analogue of the **geometric distribution**, and it has the key property of being **memoryless**. In addition to being used for the analysis of Poisson processes, it is found in various other contexts.

The exponential distribution is not the same as the class of **exponential families** of distributions, which is a large class of probability distributions that includes the exponential distribution as one of its members, but also includes the **normal distribution**, **binomial distribution**, **gamma distribution**, **Poisson**, and many others.

2.6.1 Characterization

Probability density function

The **probability density function** (pdf) of an exponential distribution is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Alternatively, this can be defined using the **Heaviside step function**, $H(x)$.

$$f(x; \lambda) = \lambda e^{-\lambda x} H(x)$$

Here $\lambda > 0$ is the parameter of the distribution, often called the *rate parameter*. The distribution is supported on the interval $[0, \infty)$. If a **random variable** X has this distribution, we write $X \sim \text{Exp}(\lambda)$.

The exponential distribution exhibits infinite divisibility.

Cumulative distribution function

The **cumulative distribution function** is given by

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Alternatively, this can be defined using the **Heaviside step function**, $H(x)$.

$$F(x; \lambda) = (1 - e^{-\lambda x})H(x)$$

Alternative parameterization

A commonly used alternative parametrization is to define the **probability density function** (pdf) of an exponential distribution as

$$f(x; \beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

where $\beta > 0$ is **mean**, **standard deviation**, and **scale parameter** of the distribution, the **reciprocal** of the **rate parameter**, λ , defined above. In this specification, β is a *survival parameter* in the sense that if a **random variable** X is the duration of time that a given biological or mechanical system manages to survive and $X \sim \text{Exp}(\beta)$ then $E[X] = \beta$. That is to say, the expected duration of survival of the system is β units of time. The parametrization involving the “rate” parameter arises in the context of events arriving at a rate λ , when the time between events (which might be modeled using an exponential distribution) has a mean of $\beta = \lambda^{-1}$.

The alternative specification is sometimes more convenient than the one given above, and some authors will use it as a standard definition. This alternative specification is not used here. Unfortunately this gives rise to a **notational** ambiguity. In general, the reader must check which of these two specifications is being used if an author writes “ $X \sim \text{Exp}(\lambda)$ ”, since either the notation in the previous (using λ) or the notation in this section (here, using β to avoid confusion) could be intended. An example of this notational switch: reference^[1] uses λ for β .

2.6.2 Properties

Mean, variance, moments and median

The mean or **expected value** of an exponentially distributed random variable X with rate parameter λ is given by

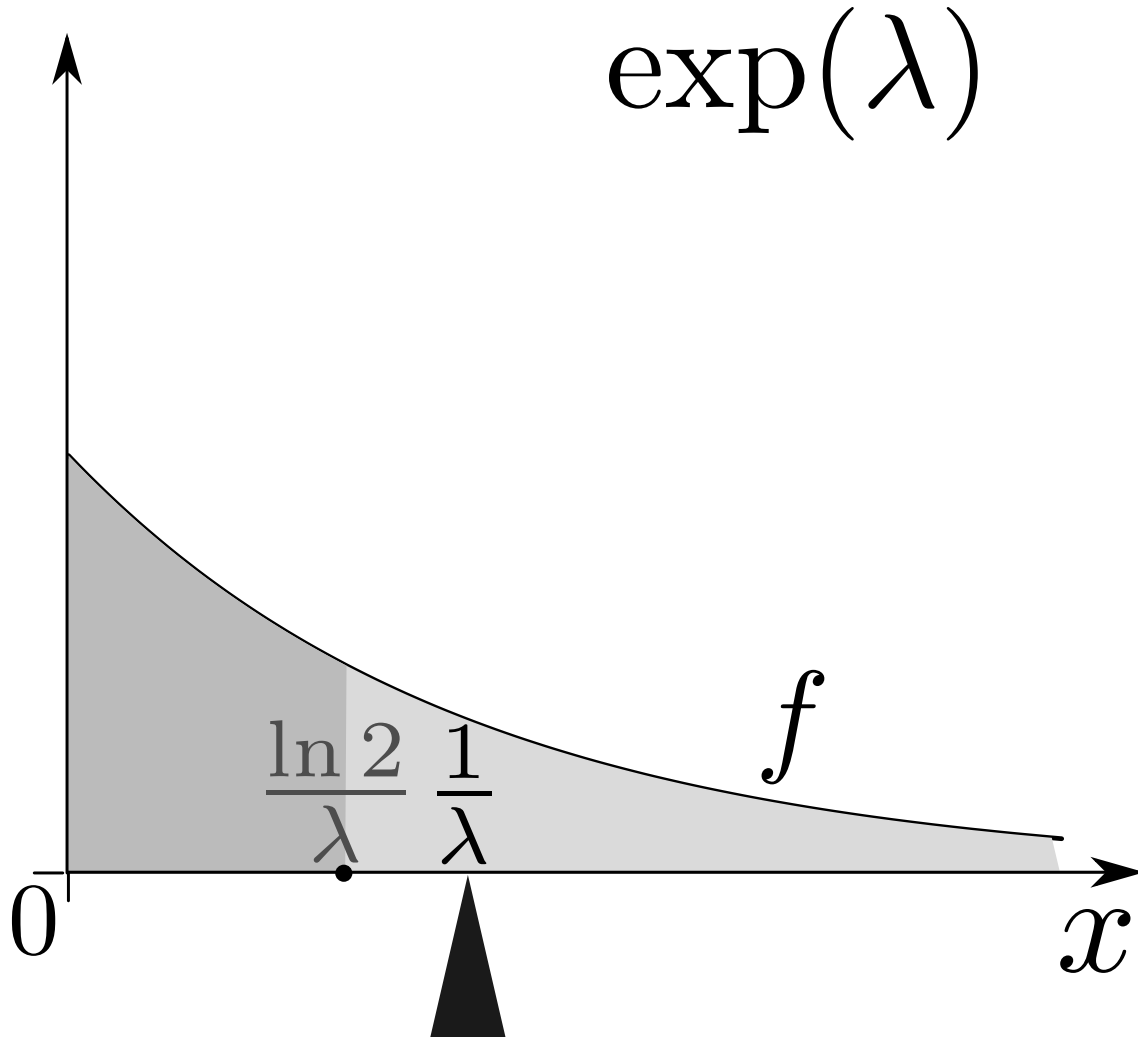
$$E[X] = \frac{1}{\lambda} = \beta$$

In light of the examples given above, this makes sense: if you receive phone calls at an average rate of 2 per hour, then you can expect to wait half an hour for every call.

The **variance** of X is given by

$$\text{Var}[X] = \frac{1}{\lambda^2}$$

so the **standard deviation** is equal to the mean.



The mean is the probability mass centre, that is the *first moment*.

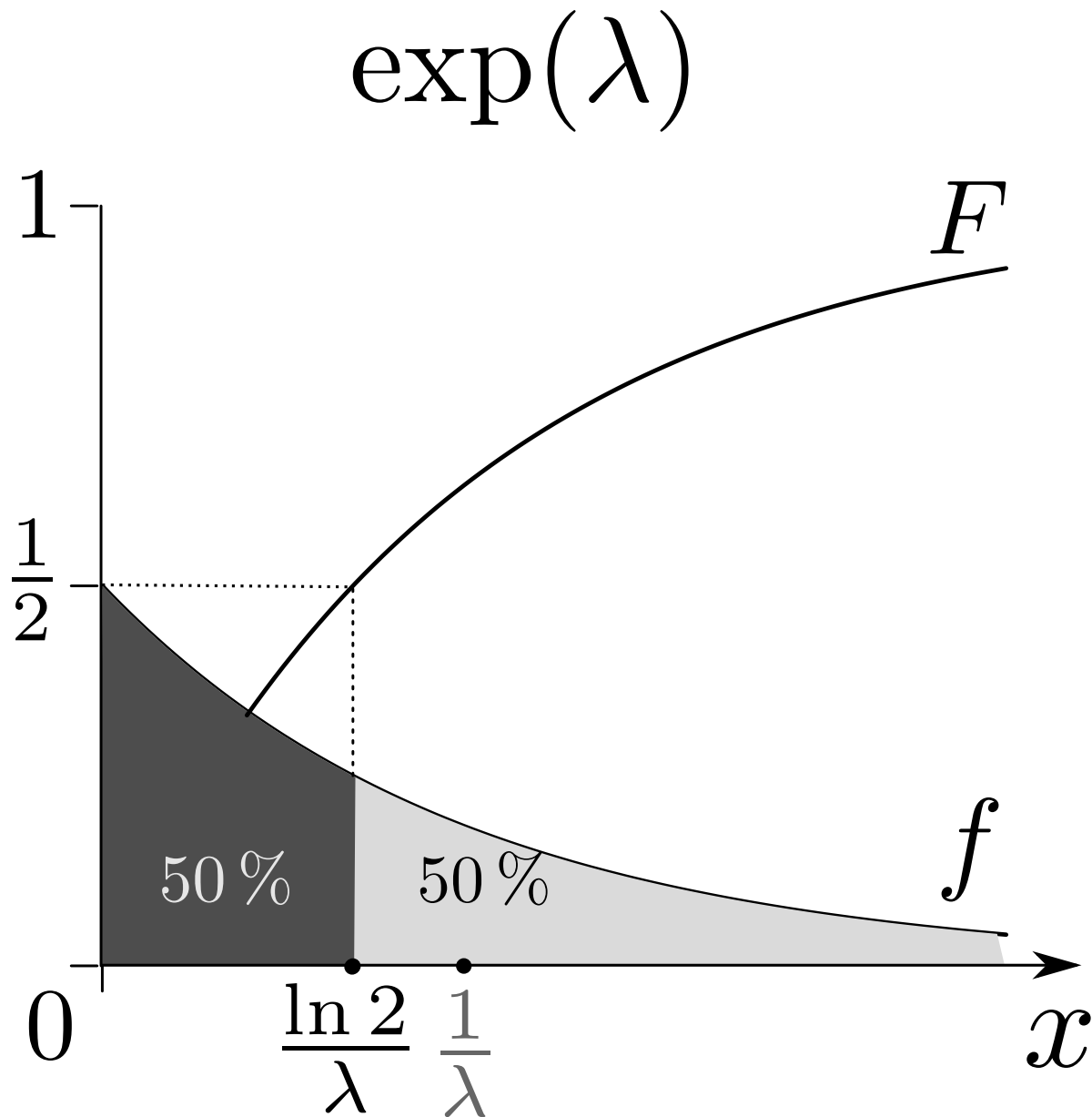
The **moments** of X , for $n = 1, 2, \dots$, are given by

$$E[X^n] = \frac{n!}{\lambda^n}$$

The **median** of X is given by

$$m[X] = \frac{\ln(2)}{\lambda} < E[X]$$

where $\ln$ refers to the natural logarithm. Thus the **absolute difference** between the mean and median is



The median is the preimage $F^{-1}(1/2)$.

$$|E[X] - m[X]| = \frac{1 - \ln(2)}{\lambda} < \frac{1}{\lambda} = \text{deviation standard}$$

in accordance with the **median-mean inequality**.

Memorylessness

An exponentially distributed random variable T obeys the relation

$$\Pr(T > s + t | T > s) = \Pr(T > t), \quad \forall s, t \geq 0$$

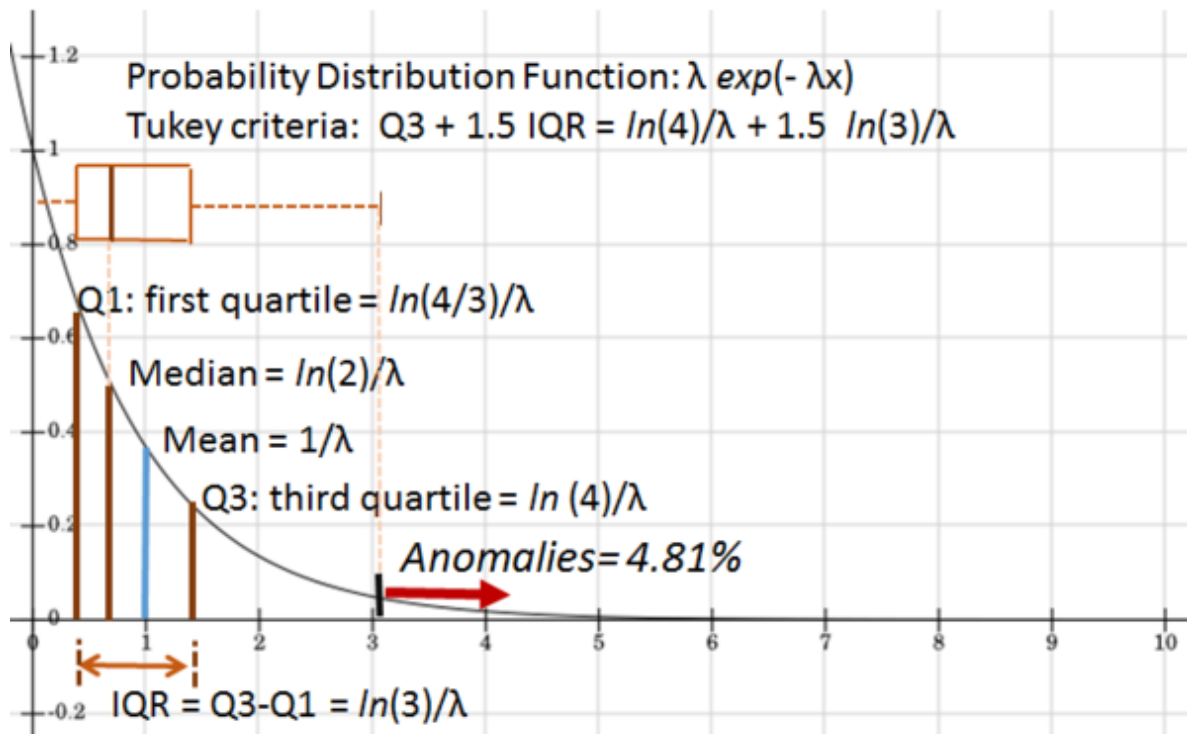
When T is interpreted as the waiting time for an event to occur relative to some initial time, this relation implies that, if T is conditioned on a failure to observe the event over some initial period of time s , the distribution of the remaining waiting time is the same as the original unconditional distribution. For example, if an event has not occurred after 30 seconds, the **conditional probability** that occurrence will take at least 10 more seconds is equal to the unconditional probability of observing the event more than 10 seconds relative to the initial time.

The exponential distribution and the **geometric distribution** are the only memoryless probability distributions.

The exponential distribution is consequently also necessarily the only continuous probability distribution that has a constant **Failure rate**.

Quantiles

Tukey criteria for anomalies. Exponential probability distribution function.



Tukey criteria for anomalies.^[2]

The quantile function (inverse cumulative distribution function) for $\text{Exp}(\lambda)$ is

$$F^{-1}(p; \lambda) = \frac{-\ln(1-p)}{\lambda}, \quad 0 \leq p < 1$$

The **quantiles** are therefore:

- first quartile: $\ln(4/3)/\lambda$

- median: $\ln(2)/\lambda$
- third quartile: $\ln(4)/\lambda$

And as a consequence the interquartile range is $\ln(3)/\lambda$.

Kullback–Leibler divergence

The directed Kullback–Leibler divergence of e^λ ('approximating' distribution) from e^{λ_0} ('true' distribution) is given by

$$\Delta(\lambda_0||\lambda) = \log(\lambda_0) - \log(\lambda) + \frac{\lambda}{\lambda_0} - 1.$$

Maximum entropy distribution

Among all continuous probability distributions with support $[0, \infty)$ and mean μ , the exponential distribution with $\lambda = \frac{1}{\mu}$ has the largest differential entropy. In other words, it is the maximum entropy probability distribution for a random variate X which is greater than or equal to zero and for which $E[X]$ is fixed.^[3]

Distribution of the minimum of exponential random variables

Let $X_1, \dots, X_n$ be independent exponentially distributed random variables with rate parameters $\lambda_1, \dots, \lambda_n$. Then

$$\min\{X_1, \dots, X_n\}$$

is also exponentially distributed, with parameter

$$\lambda = \lambda_1 + \dots + \lambda_n$$

This can be seen by considering the complementary cumulative distribution function:

$$\begin{aligned} \Pr(\min\{X_1, \dots, X_n\} > x) &= \Pr(X_1 > x, \dots, X_n > x) \\ &= \prod_{i=1}^n \Pr(X_i > x) \\ &= \prod_{i=1}^n \exp(-x\lambda_i) = \exp\left(-x \sum_{i=1}^n \lambda_i\right). \end{aligned}$$

The index of the variable which achieves the minimum is distributed according to the law

$$\Pr(X_k = \min\{X_1, \dots, X_n\}) = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_n}.$$

Note that

$$\max\{X_1, \dots, X_n\}$$

is not exponentially distributed.

2.6.3 Parameter estimation

Suppose a given variable is exponentially distributed and the rate parameter λ is to be estimated. Among the estimators of λ , the **maximum likelihood estimator** (MLE) and the **uniformly minimum variance unbiased estimator** (UMVUE) are $1/(\text{mean sample}|\bar{X})$ and $(n-1)/(\bar{X} * n)$, respectively. The one that minimizes the expected **mean squared error** is $(n-2)/(\bar{X} * n)$.^[4]

Maximum likelihood

The **likelihood function** for λ , given an **independent and identically distributed** sample $x = (x_1, \dots, x_n)$ drawn from the variable, is:

$$L(\lambda) = \prod_{i=1}^n \lambda \exp(-\lambda x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) = \lambda^n \exp(-\lambda n\bar{x}),$$

where:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the sample mean.

The derivative of the likelihood function's logarithm is:

$$\frac{d}{d\lambda} \ln(L(\lambda)) = \frac{d}{d\lambda} (n \ln(\lambda) - \lambda n\bar{x}) = \frac{n}{\lambda} - n\bar{x} \begin{cases} > 0, & 0 < \lambda < \frac{1}{\bar{x}}, \\ = 0, & \lambda = \frac{1}{\bar{x}}, \\ < 0, & \lambda > \frac{1}{\bar{x}}. \end{cases}$$

Consequently the **maximum likelihood estimate** for the rate parameter is:

$$\hat{\lambda} = \frac{1}{\bar{x}}.$$

Although this is *not* an **unbiased estimator** of λ , $\bar{x}$ is an unbiased^[5] MLE^[6] estimator of $1/\lambda = \beta$, where β is the scale parameter defined in the '**Alternative parameterization**' section above and the distribution mean.

Confidence intervals

The $100(1 - \alpha)\%$ confidence interval for the rate parameter of an exponential distribution is given by:^[7]

$$\frac{2n}{\hat{\lambda} \chi_{1-\frac{\alpha}{2}, 2n}^2} < \frac{1}{\lambda} < \frac{2n}{\hat{\lambda} \chi_{\frac{\alpha}{2}, 2n}^2}$$

which is also equal to:

$$\frac{2n\bar{x}}{\chi_{1-\frac{\alpha}{2}, 2n}^2} < \frac{1}{\lambda} < \frac{2n\bar{x}}{\chi_{\frac{\alpha}{2}, 2n}^2}$$

where $\chi^2_{p,\nu}$ is the 100(p) percentile of the chi squared distribution with ν degrees of freedom, n is the number of observations of inter-arrival times in the sample, and $\bar{x}$ is the sample average. A simple approximation to the exact interval endpoints can be derived using a normal approximation to the $\chi^2_{p,\nu}$ distribution. This approximation gives the following values for a 95% confidence interval:

$$\lambda_{low} = \hat{\lambda} \left(1 - \frac{1.96}{\sqrt{n}} \right)$$

$$\lambda_{upp} = \hat{\lambda} \left(1 + \frac{1.96}{\sqrt{n}} \right)$$

This approximation may be acceptable for samples containing at least 15 to 20 elements.^[8]

Bayesian inference

The conjugate prior for the exponential distribution is the gamma distribution (of which the exponential distribution is a special case). The following parameterization of the gamma probability density function is useful:

$$\text{Gamma}(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda\beta).$$

The posterior distribution p can then be expressed in terms of the likelihood function defined above and a gamma prior:

$$\begin{aligned} p(\lambda) &\propto L(\lambda) \times \text{Gamma}(\lambda; \alpha, \beta) \\ &= \lambda^n \exp(-\lambda n\bar{x}) \times \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda\beta) \\ &\propto \lambda^{(\alpha+n)-1} \exp(-\lambda(\beta + n\bar{x})). \end{aligned}$$

Now the posterior density p has been specified up to a missing normalizing constant. Since it has the form of a gamma pdf, this can easily be filled in, and one obtains:

$$p(\lambda) = \text{Gamma}(\lambda; \alpha + n, \beta + n\bar{x}).$$

Here the hyperparameter α can be interpreted as the number of prior observations, and β as the sum of the prior observations. The posterior mean here is:

$$\frac{\alpha + n}{\beta + n\bar{x}}.$$

2.6.4 Generating exponential variates

A conceptually very simple method for generating exponential variates is based on inverse transform sampling: Given a random variate U drawn from the uniform distribution on the unit interval $(0, 1)$, the variate

$$T = F^{-1}(U)$$

has an exponential distribution, where F^{-1} is the quantile function, defined by

$$F^{-1}(p) = \frac{-\ln(1-p)}{\lambda}.$$

Moreover, if U is uniform on $(0, 1)$, then so is $1 - U$. This means one can generate exponential variates as follows:

$$T = \frac{-\ln(U)}{\lambda}.$$

Other methods for generating exponential variates are discussed by Knuth^[9] and Devroye.^[10]

The **ziggurat algorithm** is a fast method for generating exponential variates.

A fast method for generating a set of ready-ordered exponential variates without using a sorting routine is also available.^[10]

2.6.5 Related distributions

- Exponential distribution is closed under scaling by a positive factor. If $X \sim \text{Exp}(\lambda)$ then $kX \sim \text{Exp}(\lambda/k)$.
- If $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\nu)$ then $\min(X, Y) \sim \text{Exp}(\lambda + \nu)$.
- If $X_i \sim \text{Exp}(\lambda)$ then $\min\{X_1, \dots, X_n\} \sim \text{Exp}(n\lambda)$.
- The **Benktander Weibull distribution** reduces to a truncated exponential distribution. If $X \sim \text{Exp}(\lambda)$ then $1+X \sim \text{BenktanderWeibull}(\lambda, 1)$.
- The exponential distribution is a limit of a scaled **beta distribution**:

$$\lim_{n \rightarrow \infty} n\text{Beta}(1, n) = \text{Exp}(1).$$

- If $X_i \sim \text{Exp}(\lambda)$ then the sum $X_1 + \dots + X_k = \sum_i X_i \sim \text{Erlang}(k, \lambda)$ which is just a **Gamma**(k, λ^{-1}) (in (k, θ) parametrization) or **Gamma**(k, λ) (in (α, β) parametrization) with an integer shape parameter k .
- If $X \sim \text{Exp}(1)$ then $\mu - \sigma \log(X) \sim \text{GEV}(\mu, \sigma, 0)$.
- If $X \sim \text{Exp}(\lambda)$ then $X \sim \text{Gamma}(1, \lambda^{-1})$ (in (k, θ) parametrization) or **Gamma**($1, \lambda$) (in (α, β) parametrization).
- If $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\nu)$ then $\lambda X - \nu Y \sim \text{Laplace}(0, 1)$.
- If $X, Y \sim \text{Exp}(\lambda)$ then $X - Y \sim \text{Laplace}(0, \lambda^{-1})$.
- If $X \sim \text{Laplace}(\mu, \beta^{-1})$ then $|X - \mu| \sim \text{Exp}(\beta)$.
- If $X \sim \text{Exp}(1)$ then (**logistic distribution**):

$$\mu - \beta \log \left(\frac{e^{-X}}{1 - e^{-X}} \right) \sim \text{Logistic}(\mu, \beta)$$

- If $X, Y \sim \text{Exp}(1)$ then (logistic distribution):

$$\mu - \beta \log\left(\frac{X}{Y}\right) \sim \text{Logistic}(\mu, \beta)$$

- If $X \sim \text{Exp}(\lambda)$ then $ke^X \sim \text{Pareto}(k, \lambda)$.
- If $X \sim \text{Pareto}(1, \lambda)$ then $\log(X) \sim \text{Exp}(\lambda)$.
- Exponential distribution is a special case of type 3 Pearson distribution.
- If $X \sim \text{Exp}(\lambda)$ then $\frac{e^{-X}}{k} \sim \text{PowerLaw}(k, \lambda)$ (power law)
- If $X \sim \text{Rayleigh}(\lambda^{-1/2})$ then $X^2/2 \sim \text{Exp}(\lambda)$.
- If $X \sim \text{Exp}(\lambda)$ then $X \sim \text{Weibull}(\frac{1}{\lambda}, 1)$ (Weibull distribution)
- If $X_i \sim U(0, 1)$ then

$$\lim_{n \rightarrow \infty} n \min(X_1, \dots, X_n) \sim \text{Exp}(1)$$

- If $Y|X \sim \text{Poisson}(X)$ where $X \sim \text{Exp}(\lambda^{-1})$ then $Y \sim \text{Geometric}(\frac{1}{1+\lambda})$ (geometric distribution)
- If $X \sim \text{Exp}(1)$ and $Y \sim \Gamma(\alpha, \frac{\beta}{\alpha})$ then $\sqrt{XY} \sim K(\alpha, \beta)$ (K-distribution)
- The Hoyt distribution can be obtained from Exponential distribution and Arcsine distribution
- If $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Erlang}(n, \lambda)$ then:

$$\frac{X}{Y} \sim \text{Pareto}(1, n)$$

- If $X \sim \text{Exp}(\lambda)$ and $Y \sim \Gamma(n, \frac{1}{\lambda})$ then $\frac{X}{Y} \sim \text{Pareto}(1, n)$
- If $X \sim \text{SkewLogistic}(\theta)$, then $\log(1 + e^{-X}) \sim \text{Exp}(\theta)$.
- If $X \sim \text{Exp}(\lambda)$ and $Y = \mu - \beta \log(X\lambda)$ then $Y \sim \text{Gumbel}(\mu, \beta)$.
- If $X \sim \text{Exp}(1/2)$ then $X \sim \chi_2^2$, i.e. X has a chi-squared distribution with 2 degrees of freedom.
- Let $X \sim \text{Exp}(\lambda_X)$ and $Y \sim \text{Exp}(\lambda_Y)$ be independent. Then $Z = \frac{\lambda_X X}{\lambda_Y Y}$ has probability density function $f_Z(z) = \frac{1}{(z+1)^2}$. This can be used to obtain a confidence interval for $\frac{\lambda_X}{\lambda_Y}$.

Other related distributions:

- Hyper-exponential distribution – the distribution whose density is a weighted sum of exponential densities.
- Hypoexponential distribution – the distribution of a general sum of exponential random variables.
- exGaussian distribution – the sum of an exponential distribution and a normal distribution.

2.6.6 Applications of exponential distribution

Occurrence of events

The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process.

The exponential distribution may be viewed as a continuous counterpart of the geometric distribution, which describes the number of Bernoulli trials necessary for a discrete process to change state. In contrast, the exponential distribution describes the time for a continuous process to change state.

In real-world scenarios, the assumption of a constant rate (or probability per unit time) is rarely satisfied. For example, the rate of incoming phone calls differs according to the time of day. But if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives. Similar caveats apply to the following examples which yield approximately exponentially distributed variables:

- The time until a radioactive particle decays, or the time between clicks of a geiger counter
- The time it takes before your next telephone call
- The time until default (on payment to company debt holders) in reduced form credit risk modeling

Exponential variables can also be used to model situations where certain events occur with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between roadkills on a given road.

In queuing theory, the service times of agents in a system (e.g. how long it takes for a bank teller etc. to serve a customer) are often modeled as exponentially distributed variables. (The arrival of customers for instance is also modeled by the Poisson distribution if the arrivals are independent and distributed identically.) The length of a process that can be thought of as a sequence of several independent tasks follows the Erlang distribution (which is the distribution of the sum of several independent exponentially distributed variables).

Reliability theory and reliability engineering also make extensive use of the exponential distribution. Because of the memoryless property of this distribution, it is well-suited to model the constant hazard rate portion of the bathtub curve used in reliability theory. It is also very convenient because it is so easy to add failure rates in a reliability model. The exponential distribution is however not appropriate to model the overall lifetime of organisms or technical devices, because the “failure rates” here are not constant: more failures occur for very young and for very old systems.

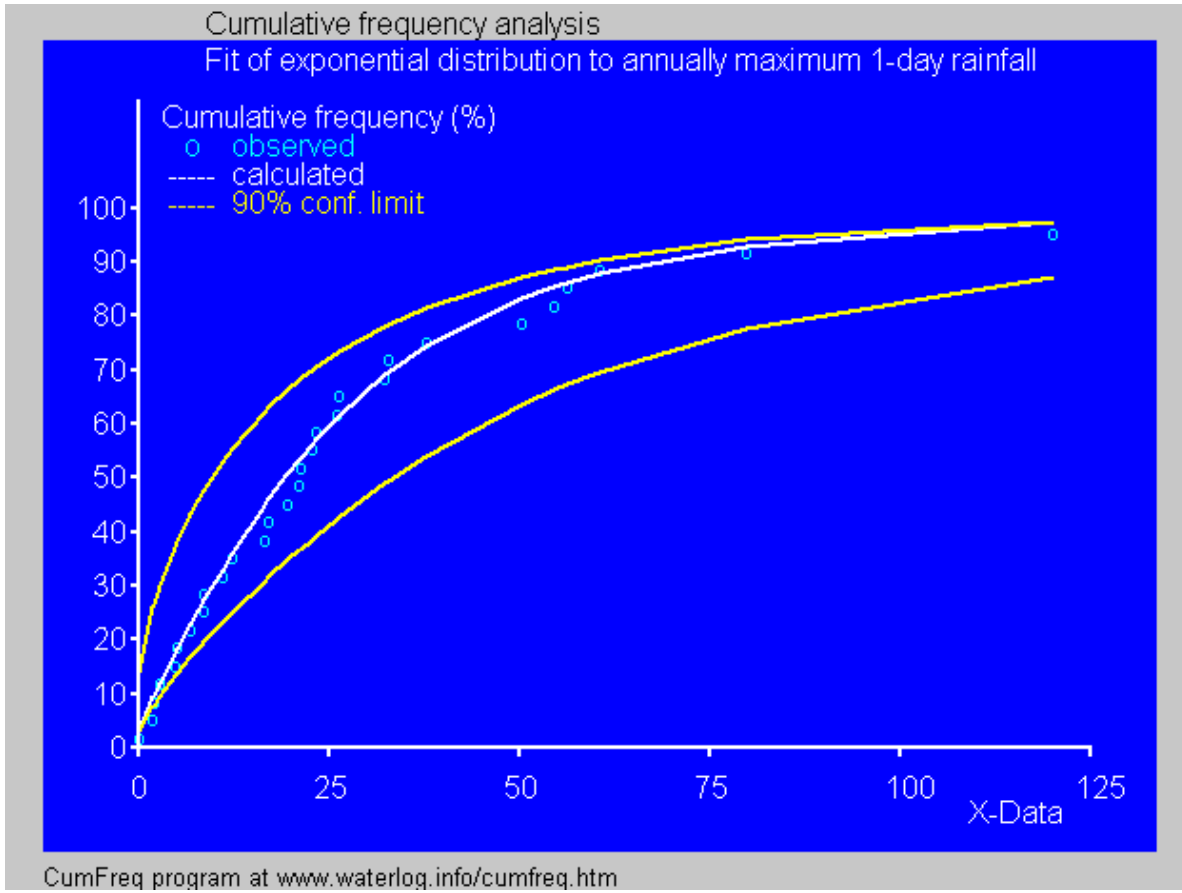
In physics, if you observe a gas at a fixed temperature and pressure in a uniform gravitational field, the heights of the various molecules also follow an approximate exponential distribution, known as the Barometric formula. This is a consequence of the entropy property mentioned below.

In hydrology, the exponential distribution is used to analyze extreme values of such variables as monthly and annual maximum values of daily rainfall and river discharge volumes.^[12]

The blue picture illustrates an example of fitting the exponential distribution to ranked annually maximum one-day rainfalls showing also the 90% confidence belt based on the binomial distribution. The rainfall data are represented by plotting positions as part of the cumulative frequency analysis.

Prediction

Having observed a sample of n data points from an unknown exponential distribution a common task is to use these samples to make predictions about future data from the same source. A common predictive distribution over future samples is the so-called plug-in distribution, formed by plugging a suitable estimate for the rate parameter λ into the exponential density function. A common choice of estimate is the one provided by the principle of maximum likelihood, and using this yields the predictive density over a future sample x_{n+1} , conditioned on the observed samples $x = (x_1, \dots, x_n)$ given by



Fitted cumulative exponential distribution to annually maximum 1-day rainfalls using CumFreq^[11]

$$p_{ML}(x_{n+1} \mid x_1, \dots, x_n) = \left(\frac{1}{\bar{x}}\right) \exp\left(-\frac{x_{n+1}}{\bar{x}}\right)$$

The Bayesian approach provides a predictive distribution which takes into account the uncertainty of the estimated parameter, although this may depend crucially on the choice of prior.

A predictive distribution free of the issues of choosing priors that arise under the subjective Bayesian approach is

$$p_{CNML}(x_{n+1} \mid x_1, \dots, x_n) = \frac{n^{n+1} (\bar{x})^n}{(n\bar{x} + x_{n+1})^{n+1}},$$

which can be considered as

- (1) a frequentist **confidence distribution**, obtained from the distribution of the pivotal quantity $x_{n+1}/\bar{x}$; ^[13]
- (2) a profile predictive likelihood, obtained by eliminating the parameter λ from the joint likelihood of x_{n+1} and λ by maximization; ^[14]
- (3) an objective Bayesian predictive posterior distribution, obtained using the non-informative **Jeffreys prior** $1/\lambda$;
- (4) the Conditional Normalized Maximum Likelihood (CNML) predictive distribution, from information theoretic considerations. ^[15]

The accuracy of a predictive distribution may be measured using the distance or divergence between the true exponential distribution with rate parameter, λ_0 , and the predictive distribution based on the sample x . The **Kullback–Leibler divergence** is a commonly used, parameterisation free measure of the difference between two distributions. Letting $\Delta(\lambda_0 \| p)$ denote the Kullback–Leibler divergence between an exponential with rate parameter λ_0 and a predictive distribution p it can be shown that

$$E_{\lambda_0} [\Delta(\lambda_0 \| p_{\text{ML}})] = \psi(n) + \frac{1}{n-1} - \log(n)$$

$$E_{\lambda_0} [\Delta(\lambda_0 \| p_{\text{CNML}})] = \psi(n) + \frac{1}{n} - \log(n)$$

where the expectation is taken with respect to the exponential distribution with rate parameter $\lambda_0 \in (0, \infty)$, and $\psi(\cdot)$ is the digamma function. It is clear that the CNML predictive distribution is strictly superior to the maximum likelihood plug-in distribution in terms of average Kullback–Leibler divergence for all sample sizes $n > 0$.

2.6.7 See also

- **Dead time** – an application of exponential distribution to particle detector analysis.
- **Laplace distribution**, or the “double exponential distribution”.

2.6.8 References

- [1] David Olive, *Chapter 4. Truncated Distributions*, “Lemma 4.3”, Southern Illinois University, February 18, 2010, p.107.
- [2] Brillinger, David R. (2011). “Data analysis, exploratory” (PDF). SAGE Publications. pp. 530–537. doi:10.4135/9781412959636.n128. Retrieved 2013-11-21.
- [3] Park, Sung Y.; Bera, Anil K. (2009). “Maximum entropy autoregressive conditional heteroskedasticity model” (PDF). *Journal of Econometrics* (Elsevier): 219–230. Retrieved 2011-06-02.
- [4] Abdulaziz Elfessi and David M. Reineke, “A Bayesian Look at Classical Estimation: The Exponential Distribution”, *Journal of Statistics Education* Volume 9, Number 1 (2001).
- [5] Richard Arnold Johnson; Dean W. Wichern (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall. ISBN 978-0-13-187715-3. Retrieved 10 August 2012.
- [6] *NIST/SEMATECH e-Handbook of Statistical Methods*
- [7] Ross, Sheldon M. (2009). *Introduction to probability and statistics for engineers and scientists* (4th ed.). Associated Press. p. 267. ISBN 978-0-12-370483-2.
- [8] Guerriero, V. (2012). “Power Law Distribution: Method of Multi-scale Inferential Statistics”. *Journal of Modern Mathematics Frontier (JMMF)* **1**: 21–28.
- [9] Donald E. Knuth (1998). *The Art of Computer Programming*, volume 2: *Seminumerical Algorithms*, 3rd edn. Boston: Addison–Wesley. ISBN 0-201-89684-2. See section 3.4.1, p. 133.
- [10] Luc Devroye (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag. ISBN 0-387-96305-7. See chapter IX, section 2, pp. 392–401.
- [11] “Cumfreq, a free computer program for cumulative frequency analysis”.
- [12] Ritzema (ed.), H.P. (1994). *Frequency and Regression Analysis* (PDF). Chapter 6 in: Drainage Principles and Applications, Publication 16, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. pp. 175–224. ISBN 90-70754-33-9.
- [13] Lawless, J.F., Fredette, M., “Frequentist predictions intervals and predictive distributions”, *Biometrika* (2005), Vol 92, Issue 3, pp 529–542.

- [14] Bjornstad, J.F. (1990). "Predictive Likelihood: A Review". *Statist. Sci.* **5** (2): 242–254. doi:10.1214/ss/1177012175.
- [15] D. F. Schmidt and E. Makalic, "Universal Models for the Exponential Distribution", *IEEE Transactions on Information Theory*, Volume 55, Number 7, pp. 3087–3090, 2009 doi:10.1109/TIT.2009.2018331

2.6.9 External links

- Hazewinkel, Michiel, ed. (2001), "Exponential distribution", *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Online calculator of Exponential Distribution

2.7 F-distribution

Not to be confused with *F*-statistics as used in population genetics.

The **F-distribution**, also known as **Snedecor's F distribution** or the **Fisher–Snedecor distribution** (after Ronald Fisher and George W. Snedecor) is, in probability theory and statistics, a continuous probability distribution.^{[1][2][3][4]} The *F*-distribution arises frequently as the null distribution of a test statistic, most notably in the analysis of variance; see *F*-test.

2.7.1 Definition

If a random variable *X* has an *F*-distribution with parameters *d*₁ and *d*₂, we write *X* ~ F(*d*₁, *d*₂). Then the probability density function (pdf) for *X* is given by

$$\begin{aligned} f(x; d_1, d_2) &= \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x \operatorname{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \\ &= \frac{1}{\operatorname{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2} - 1} \left(1 + \frac{d_1}{d_2} x\right)^{-\frac{d_1 + d_2}{2}} \end{aligned}$$

for real $x \geq 0$. Here B is the **beta function**. In many applications, the parameters *d*₁ and *d*₂ are positive integers, but the distribution is well-defined for positive real values of these parameters.

The cumulative distribution function is

$$F(x; d_1, d_2) = I_{\frac{d_1 x}{d_1 x + d_2}}\left(\frac{d_1}{2}, \frac{d_2}{2}\right),$$

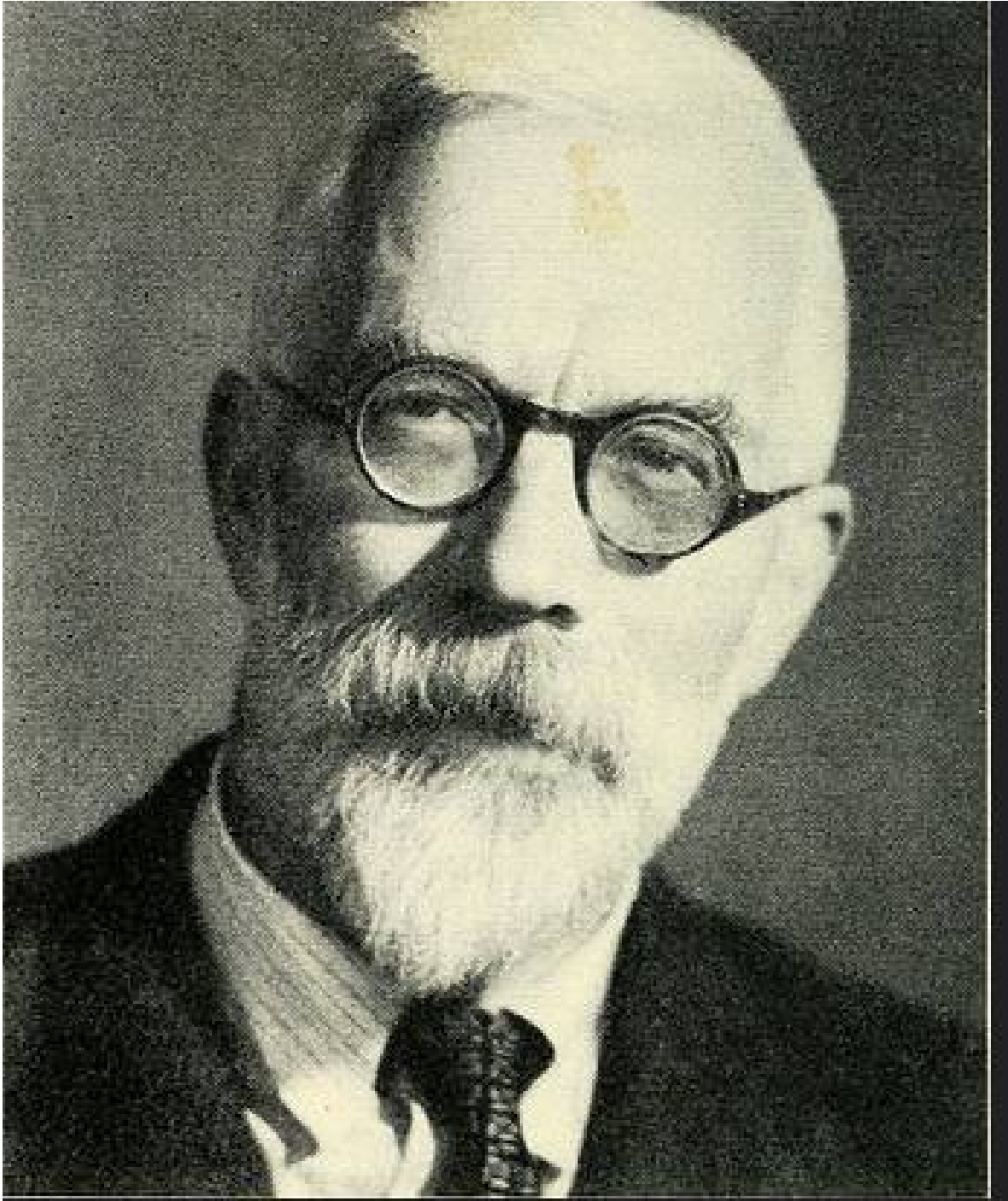
where *I* is the **regularized incomplete beta function**.

The expectation, variance, and other details about the F(*d*₁, *d*₂) are given in the sidebox; for *d*₂ > 8, the excess kurtosis is

$$\gamma_2 = 12 \frac{d_1(5d_2 - 22)(d_1 + d_2 - 2) + (d_2 - 4)(d_2 - 2)^2}{d_1(d_2 - 6)(d_2 - 8)(d_1 + d_2 - 2)}$$

The *k*-th moment of an F(*d*₁, *d*₂) distribution exists and is finite only when $2k < d_2$ and it is equal to ^[5]

$$\mu_X(k) = \left(\frac{d_2}{d_1}\right)^k \frac{\Gamma\left(\frac{d_1}{2} + k\right)}{\Gamma\left(\frac{d_1}{2}\right)} \frac{\Gamma\left(\frac{d_2}{2} - k\right)}{\Gamma\left(\frac{d_2}{2}\right)}$$



Biologist and statistician Ronald Fisher

The F -distribution is a particular parametrization of the **beta prime distribution**, which is also called the beta distribution of the second kind.

The **characteristic function** is listed incorrectly in many standard references (e.g., ^[2]). The correct expression ^[6] is

$$\varphi_{d_1, d_2}^F(s) = \frac{\Gamma(\frac{d_1+d_2}{2})}{\Gamma(\frac{d_2}{2})} U\left(\frac{d_1}{2}, 1 - \frac{d_2}{2}, -\frac{d_2}{d_1} s\right)$$

where $U(a, b, z)$ is the confluent hypergeometric function of the second kind.

2.7.2 Characterization

A random variate of the F -distribution with parameters d_1 and d_2 arises as the ratio of two appropriately scaled chi-squared variates:^[7]

$$X = \frac{U_1/d_1}{U_2/d_2}$$

where

- U_1 and U_2 have chi-squared distributions with d_1 and d_2 degrees of freedom respectively, and
- U_1 and U_2 are independent.

In instances where the F -distribution is used, for example in the analysis of variance, independence of U_1 and U_2 might be demonstrated by applying Cochran's theorem.

Equivalently, the random variable of the F -distribution may also be written

$$X = \frac{s_1^2}{\sigma_1^2} / \frac{s_2^2}{\sigma_2^2}$$

where s_1^2 and s_2^2 are the sums of squares S_1^2 and S_2^2 from two normal processes with variances σ_1^2 and σ_2^2 divided by the corresponding number of χ^2 degrees of freedom, d_1 and d_2 respectively.

In a frequentist context, a scaled F -distribution therefore gives the probability $p(s_1^2/s_2^2 \mid \sigma_1^2, \sigma_2^2)$, with the F -distribution itself, without any scaling, applying where σ_1^2 is being taken equal to σ_2^2 . This is the context in which the F -distribution most generally appears in **F -tests**: where the null hypothesis is that two independent normal variances are equal, and the observed sums of some appropriately selected squares are then examined to see whether their ratio is significantly incompatible with this null hypothesis.

The quantity X has the same distribution in Bayesian statistics, if an uninformative rescaling-invariant Jeffreys prior is taken for the prior probabilities of σ_1^2 and σ_2^2 .^[8] In this context, a scaled F -distribution thus gives the posterior probability $p(\sigma_2^2/\sigma_1^2 \mid s_1^2, s_2^2)$, where now the observed sums s_1^2 and s_2^2 are what are taken as known.

Differential equation

The probability density function of the F -distribution is a solution of the following differential equation:

$$\left\{ \begin{array}{l} 2x(d_1x + d_2)f'(x) + (2d_1x + d_2d_1x - d_2d_1 + 2d_2)f(x) = 0, \\ f(1) = \frac{d_1^{\frac{d_1}{2}} d_2^{\frac{d_2}{2}} (d_1 + d_2)^{\frac{1}{2}(-d_1 - d_2)}}{B(\frac{d_1}{2}, \frac{d_2}{2})} \end{array} \right\}$$

2.7.3 Generalization

A generalization of the (central) F -distribution is the noncentral F -distribution.

2.7.4 Related distributions and properties

- If $X \sim \chi_{d_1}^2$ and $Y \sim \chi_{d_2}^2$ are **independent**, then $\frac{X/d_1}{Y/d_2} \sim F(d_1, d_2)$
- If $X \sim \text{Beta}(d_1/2, d_2/2)$ (**Beta distribution**) then $\frac{d_2 X}{d_1(1-X)} \sim F(d_1, d_2)$
- Equivalently, if $X \sim F(d_1, d_2)$, then $\frac{d_1 X/d_2}{1+d_1 X/d_2} \sim \text{Beta}(d_1/2, d_2/2)$.
- If $X \sim F(d_1, d_2)$ then $Y = \lim_{d_2 \rightarrow \infty} d_1 X$ has the **chi-squared distribution** $\chi_{d_1}^2$
- $F(d_1, d_2)$ is equivalent to the scaled **Hotelling's T-squared distribution** $\frac{d_2}{d_1(d_1+d_2-1)} T^2(d_1, d_1 + d_2 - 1)$.
- If $X \sim F(d_1, d_2)$ then $X^{-1} \sim F(d_2, d_1)$.
- If $X \sim t(n)$ then

$$X^2 \sim F(1, n)$$

$$X^{-2} \sim F(n, 1)$$

- F -distribution is a special case of type 6 **Pearson distribution**
- If X and Y are independent, with $X, Y \sim \text{Laplace}(\mu, b)$ then

$$\frac{|X-\mu|}{|Y-\mu|} \sim F(2, 2)$$

- If $X \sim F(n, m)$ then $\frac{\log X}{2} \sim \text{FisherZ}(n, m)$ (**Fisher's z-distribution**)
- The **noncentral F-distribution** simplifies to the F -distribution if $\lambda = 0$.
- The doubly **noncentral F-distribution** simplifies to the F -distribution if $\lambda_1 = \lambda_2 = 0$
- If $Q_X(p)$ is the quantile p for $X \sim F(d_1, d_2)$ and $Q_Y(1-p)$ is the quantile $1-p$ for $Y \sim F(d_2, d_1)$, then

$$Q_X(p) = \frac{1}{Q_Y(1-p)}$$

2.7.5 See also

- Chi-squared distribution
- Chow test
- Gamma distribution
- Hotelling's T-squared distribution
- Student's t-distribution
- Wilks' lambda distribution
- Wishart distribution

2.7.6 References

- [1] Johnson, Norman Lloyd; Samuel Kotz; N. Balakrishnan (1995). *Continuous Univariate Distributions, Volume 2 (Second Edition, Section 27)*. Wiley. ISBN 0-471-58494-0.
- [2] Abramowitz, Milton; Stegun, Irene A., eds. (December 1972) [1964]. "Chapter 26". *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Applied Mathematics Series **55** (10 ed.). New York, USA: United States Department of Commerce, National Bureau of Standards; Dover Publications. p. 946. ISBN 978-0-486-61272-0. LCCN 64-60036. MR 0167642.
- [3] NIST (2006). *Engineering Statistics Handbook – F Distribution*
- [4] Mood, Alexander; Franklin A. Graybill; Duane C. Boes (1974). *Introduction to the Theory of Statistics (Third Edition, pp. 246–249)*. McGraw-Hill. ISBN 0-07-042864-6.
- [5] Taboga, Marco. "The F distribution".
- [6] Phillips, P. C. B. (1982) "The true characteristic function of the F distribution," *Biometrika*, 69: 261–264 JSTOR 2335882
- [7] M.H. DeGroot (1986), *Probability and Statistics* (2nd Ed), Addison-Wesley. ISBN 0-201-11366-X, p. 500
- [8] G.E.P. Box and G.C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley. p.110

2.7.7 External links

- Table of critical values of the F -distribution
- Earliest Uses of Some of the Words of Mathematics: entry on F -distribution contains a brief history
- Free calculator for F -testing

2.8 Fisher's z-distribution

Not to be confused with Fisher z-transformation.

Fisher's z-distribution is the statistical distribution of half the logarithm of an F distribution variate:

$$z = \frac{1}{2} \log F$$

It was first described by Ronald Fisher in a paper delivered at the International Mathematical Congress of 1924 in Toronto, titled "On a distribution yielding the error functions of several well-known statistics" (Proceedings of the International Congress of Mathematics, Toronto, 2: 805-813 (1924). Nowadays one usually uses the F distribution instead.

The probability density function and cumulative distribution function can be found by using the F-distribution at the value of $x' = e^{2x}$. However, the mean and variance do not follow the same transformation.

The probability density function is^{[1][2]}

$$f(x; d_1, d_2) = \frac{2d_1^{d_1/2} d_2^{d_2/2}}{B(d_1/2, d_2/2)} \frac{e^{d_1 x}}{(d_1 e^{2x} + d_2)^{(d_1+d_2)/2}},$$

where B is the beta function.

When the degrees of freedom becomes large ($d_1, d_2 \rightarrow \infty$) the distribution approach normality with mean^[1]

$$\bar{x} = (1/d_2 - 1/d_1)/2$$

and variance

$$\sigma_x^2 = (1/d_1 + 1/d_2)/2.$$

2.8.1 Related Distribution

- If $X \sim \text{FisherZ}(n, m)$ then $e^{2X} \sim F(n, m)$ (F-distribution)
- If $X \sim F(n, m)$ then $\frac{\log X}{2} \sim \text{FisherZ}(n, m)$

2.8.2 References

- Fisher, R.A. (1924) On a Distribution Yielding the Error Functions of Several Well Known Statistics *Proceedings of the International Congress of Mathematics, Toronto*, 2: 805-813 [pdf copy](#)

[1] Leo A. Aroian (December 1941). "A study of R. A. Fisher's z distribution and the related F distribution". *The Annals of Mathematical Statistics* **12** (4): 429–448. doi:10.1214/aoms/1177731681. JSTOR 2235955.

[2] Charles Ernest Weatherburn. *A first course in mathematical statistics*.

2.8.3 External links

- [MathWorld entry](#)

2.9 Folded normal distribution

The **folded normal distribution** is a [probability distribution](#) related to the [normal distribution](#). Given a normally distributed random variable X with [mean](#) μ and [variance](#) σ^2 , the [random variable](#) $Y = |X|$ has a folded normal distribution. Such a case may be encountered if only the magnitude of some variable is recorded, but not its sign. The distribution is called Folded because probability mass to the left of the $x = 0$ is "folded" over by taking the [absolute value](#). In the physics of [heat conduction](#), the folded normal distribution is a fundamental solution of the [heat equation](#) on the upper plane (i.e. a [heat kernel](#)).

The [probability density function](#) (PDF) is given by

$$f_Y(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x+\mu)^2}{2\sigma^2}}$$

for $x \geq 0$, and 0 everywhere else. It follows that the [cumulative distribution function](#) (CDF) is given by:

$$F_Y(x; \mu, \sigma) = \frac{1}{2} \left[\operatorname{erf} \left(\frac{x+\mu}{\sigma\sqrt{2}} \right) + \operatorname{erf} \left(\frac{x-\mu}{\sigma\sqrt{2}} \right) \right]$$

for $x \geq 0$, where $\operatorname{erf}()$ is the [error function](#). This expression reduces to the CDF of the [half-normal distribution](#) when $\mu = 0$.

The mean of the folded distribution is then

$$\mu_Y = \sigma \sqrt{\frac{2}{\pi}} \exp \left(\frac{-\mu^2}{2\sigma^2} \right) - \mu \operatorname{erf} \left(\frac{-\mu}{\sqrt{2}\sigma} \right),$$

The variance then is expressed easily in terms of the mean:

$$\sigma_Y^2 = \mu^2 + \sigma^2 - \mu_Y^2.$$

Both the mean (μ) and variance (σ^2) of X in the original normal distribution can be interpreted as the location and scale parameters of Y in the folded distribution.

2.9.1 Differential equations

The PDF of the folded normal distribution can also be defined by the system of differential equations

$$\begin{cases} \sigma^4 f''(x) + 2\sigma^2 x f'(x) + (-\mu^2 + \sigma^2 + x^2) f(x) = 0 \\ f(0) = \sqrt{2/\pi} \frac{1}{\sigma} e^{-\frac{\mu^2}{2\sigma^2}} \\ f'(0) = 0 \end{cases}$$

2.9.2 Related distributions

- When $\mu = 0$, the distribution of Y is a half-normal distribution.
- The random variable $(Y/\sigma)^2$ has a noncentral chi-squared distribution with 1 degree of freedom and noncentrality equal to $(\mu/\sigma)^2$.

2.9.3 See also

- Folded cumulative distribution

2.9.4 External links

- Virtual Laboratories: The Folded Normal Distribution

2.9.5 References

- Leone FC, Nottingham RB, Nelson LS (1961). "The Folded Normal Distribution". *Technometrics* (Technometrics, Vol. 3, No. 4) **3** (4): 543–550. doi:10.2307/1266560. JSTOR 1266560.
- Johnson NL (1962). "The folded normal distribution: accuracy of the estimation by maximum likelihood". *Technometrics* (Technometrics, Vol. 4, No. 2) **4** (2): 249–256. doi:10.2307/1266622. JSTOR 1266622.
- Nelson LS (1980). "The Folded Normal Distribution". *J Qual Technol* **12** (4): 236–238.
- Elandt RC (1961). "The folded normal distribution: two methods of estimating parameters from moments". *Technometrics* (Technometrics, Vol. 3, No. 4) **3** (4): 551–562. doi:10.2307/1266561. JSTOR 1266561.
- Lin PC (2005). "Application of the generalized folded-normal distribution to the process capability measures". *Int J Adv Manuf Technol* **26** (7–8): 825–830. doi:10.1007/s00170-003-2043-x.

2.10 Fréchet distribution

The **Fréchet distribution**, also known as inverse Weibull distribution,^{[2][3]} is a special case of the **generalized extreme value distribution**. It has the cumulative distribution function

$$\Pr(X \leq x) = e^{-x^{-\alpha}} \text{ if } x > 0.$$

where $\alpha > 0$ is a **shape parameter**. It can be generalised to include a **location parameter** m (the minimum) and a **scale parameter** $s > 0$ with the cumulative distribution function

$$\Pr(X \leq x) = e^{-\left(\frac{x-m}{s}\right)^{-\alpha}} \text{ if } x > m.$$

Named for **Maurice Fréchet** who wrote a related paper in 1927, further work was done by **Fisher and Tippett** in 1928 and by **Gumbel** in 1958.

2.10.1 Characteristics

The single parameter Fréchet with parameter α has **standardized moment**

$$\mu_k = \int_0^\infty x^k f(x) dx = \int_0^\infty t^{-\frac{k}{\alpha}} e^{-t} dt,$$

(with $t = x^{-\alpha}$) defined only for $k < \alpha$:

$$\mu_k = \Gamma\left(1 - \frac{k}{\alpha}\right)$$

where $\Gamma(z)$ is the **Gamma function**.

In particular:

- For $\alpha > 1$ the **expectation** is $E[X] = \Gamma(1 - \frac{1}{\alpha})$
- For $\alpha > 2$ the **variance** is $\text{Var}(X) = \Gamma(1 - \frac{2}{\alpha}) - (\Gamma(1 - \frac{1}{\alpha}))^2$.

The **quantile** q_y of order y can be expressed through the inverse of the distribution,

$$q_y = F^{-1}(y) = (-\log_e y)^{-\frac{1}{\alpha}}$$

In particular the median is:

$$q_{1/2} = (\log_e 2)^{-\frac{1}{\alpha}}.$$

The **mode** of the distribution is $\left(\frac{\alpha}{\alpha+1}\right)^{\frac{1}{\alpha}}$.

Especially for the 3-parameter Fréchet, the first quartile is $q_1 = m + \frac{s}{\sqrt[\alpha]{\log(4)}}$ and the third quartile $q_3 = m + \frac{s}{\sqrt[\alpha]{\log(\frac{4}{3})}}$.

Also the quantiles for the mean and mode are:

$$F(\text{mean}) = \exp\left(-\Gamma^{-\alpha}\left(1 - \frac{1}{\alpha}\right)\right)$$

$$F(\text{mode}) = \exp\left(-\frac{\alpha + 1}{\alpha}\right).$$

2.10.2 Applications

- In **hydrology**, the Fréchet distribution is applied to extreme events such as annually maximum one-day rainfalls and river discharges.^[4] The blue picture illustrates an example of fitting the Fréchet distribution to ranked annually maximum one-day rainfalls in Oman showing also the 90% **confidence belt** based on the **binomial distribution**. The cumulative frequencies of the rainfall data are represented by **plotting positions** as part of the **cumulative frequency analysis**. However, in most hydrological applications, the distribution fitting is via the **generalized extreme value distribution** as this avoids imposing the assumption that the distribution does not have a lower bound (as required by the Frechet distribution).
- One useful test to assess whether a multivariate distribution is asymptotically dependent or independent consists transforming the data into standard Frechet margins using transformation $Z_i = -1/\log F_i(X_i)$ and then mapping from the cartesian to pseudo-polar coordinates $(R, W) = (Z_1 + Z_2, Z_1/(Z_1 + Z_2))$. $R \gg 1$ corresponds to the extreme data for which at least only one component is large while W approximately 1 or 0 corresponds to only one component being extreme.

2.10.3 Related distributions

- If $X \sim U(0, 1)$ (**Uniform distribution (continuous)**) then $m + s(-\log(X))^{-1/\alpha} \sim \text{Frechet}(\alpha, s, m)$
- If $X \sim \text{Frechet}(\alpha, s, m)$ then $kX + b \sim \text{Frechet}(\alpha, ks, km + b)$
- If $X_i \sim \text{Frechet}(\alpha, s, m)$ and $Y = \max\{X_1, \dots, X_n\}$ then $Y \sim \text{Frechet}(\alpha, n^{\frac{1}{\alpha}}s, m)$
- The **cumulative distribution function** of the Frechet distribution solves the maximum **stability postulate** equation
- If $X \sim \text{Frechet}(\alpha, s, m = 0)$ then its reciprocal is **Weibull-distributed**: $X^{-1} \sim \text{Weibull}(k = \alpha, \lambda = s^{-1})$

2.10.4 Properties

- The Frechet distribution is a **max stable distribution**
- The negative of a random variable having a Frechet distribution is a **min stable distribution**

2.10.5 See also

- **Type-2 Gumbel distribution**
- **Fisher–Tippett–Gnedenko theorem**
- **CumFreq** (application software for probability distributions including Fréchet)

2.10.6 References

- [1] Muraleedharan. G, C. Guedes Soares and Cláudia Lucas (2011). “Characteristic and Moment Generating Functions of Generalised Extreme Value Distribution (GEV)”. In Linda. L. Wright (Ed.), *Sea Level Rise, Coastal Engineering, Shorelines and Tides*, Chapter 14, pp. 269–276. Nova Science Publishers. ISBN 978-1-61728-655-1
- [2] Khan M.S., Pasha G.R. and Pasha A.H. (February 2008). “Theoretical Analysis of Inverse Weibull Distribution” (PDF). *WSEAS TRANSACTIONS on MATHEMATICS* 7 (2). pp. 30–38.
- [3] de Gusmão, Felipe R.S. and Ortega, Edwin M.M. and Cordeiro, Gauss M. (2011). “The generalized inverse Weibull distribution”. *Statistical Papers* 52 (3) (Springer-Verlag). pp. 591–619. doi:10.1007/s00362-009-0271-3. ISSN 0932-5026.
- [4] Coles, Stuart (2001). *An Introduction to Statistical Modeling of Extreme Values*,. Springer-Verlag. ISBN 1-85233-459-2.

2.10.7 Publications

- Fréchet, M., (1927). “Sur la loi de probabilité de l'écart maximum.” *Ann. Soc. Polon. Math.* 6, 93.
- Fisher, R.A., Tippett, L.H.C., (1928). “Limiting forms of the frequency distribution of the largest and smallest member of a sample.” *Proc. Cambridge Philosophical Society* 24:180–190.
- Gumbel, E.J. (1958). “Statistics of Extremes.” Columbia University Press, New York.
- Kotz, S.; Nadarajah, S. (2000) *Extreme value distributions: theory and applications*, World Scientific. ISBN 1-86094-224-5

2.10.8 External links

- Bank of England working paper
- An application of a new extreme value distribution to air pollution data
- Wave Analysis for Fatigue and Oceanography

2.11 Gamma distribution

Not to be confused with [Gamma function](#).

In probability theory and statistics, the **gamma distribution** is a two-parameter family of continuous probability distributions. The common exponential distribution and chi-squared distribution are special cases of the gamma distribution. There are three different parametrizations in common use:

1. With a shape parameter k and a scale parameter θ .
2. With a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, called a **rate parameter**.
3. With a shape parameter k and a mean parameter $\mu = k/\beta$.

In each of these three forms, both parameters are positive real numbers.

The parameterization with k and θ appears to be more common in [econometrics](#) and certain other applied fields, where e.g. the gamma distribution is frequently used to model waiting times. For instance, in [life testing](#), the waiting time until death is a [random variable](#) that is frequently modeled with a gamma distribution.^[2]

The parameterization with α and β is more common in [Bayesian statistics](#), where the gamma distribution is used as a [conjugate prior distribution](#) for various types of inverse scale (aka rate) parameters, such as the λ of an [exponential](#)

distribution or a Poisson distribution^[3] – or for that matter, the β of the gamma distribution itself. (The closely related inverse gamma distribution is used as a conjugate prior for scale parameters, such as the variance of a normal distribution.)

If k is an integer, then the distribution represents an Erlang distribution; i.e., the sum of k independent exponentially distributed random variables, each of which has a mean of θ (which is equivalent to a rate parameter of $1/\theta$).

The gamma distribution is the maximum entropy probability distribution for a random variable X for which $\mathbf{E}[X] = k\theta = \alpha/\beta$ is fixed and greater than zero, and $\mathbf{E}[\ln(X)] = \psi(k) + \ln(\theta) = \psi(\alpha) - \ln(\beta)$ is fixed (ψ is the digamma function).^[4]

2.11.1 Characterization using shape k and scale θ

A random variable X that is gamma-distributed with shape k and scale θ is denoted by

$$X \sim \Gamma(k, \theta) \equiv \text{Gamma}(k, \theta)$$

Probability density function

The probability density function using the shape-scale parametrization is

$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \quad \text{for } x > 0 \text{ and } k, \theta > 0.$$

Here $\Gamma(k)$ is the gamma function evaluated at k .

Cumulative distribution function

The cumulative distribution function is the regularized gamma function:

$$F(x; k, \theta) = \int_0^x f(u; k, \theta) du = \frac{\gamma(k, \frac{x}{\theta})}{\Gamma(k)}$$

where $\gamma(k, x/\theta)$ is the lower incomplete gamma function.

It can also be expressed as follows, if k is a positive integer (i.e., the distribution is an Erlang distribution):^[5]

$$F(x; k, \theta) = 1 - \sum_{i=0}^{k-1} \frac{1}{i!} \left(\frac{x}{\theta}\right)^i e^{-\frac{x}{\theta}} = e^{-\frac{x}{\theta}} \sum_{i=k}^{\infty} \frac{1}{i!} \left(\frac{x}{\theta}\right)^i$$

2.11.2 Characterization using shape α and rate β

Alternatively, the gamma distribution can be parameterized in terms of a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, called a rate parameter. A random variable X that is gamma-distributed with shape α and rate β is denoted

$$X \sim \Gamma(\alpha, \beta) \equiv \text{Gamma}(\alpha, \beta)$$

Probability density function

The corresponding density function in the shape-rate parametrization is

$$g(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-x\beta}}{\Gamma(\alpha)} \quad \text{for } x \geq 0 \text{ and } \alpha, \beta > 0$$

Both parametrizations are common because either can be more convenient depending on the situation.

Cumulative distribution function

The **cumulative distribution function** is the regularized gamma function:

$$F(x; \alpha, \beta) = \int_0^x f(u; \alpha, \beta) du = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

where $\gamma(\alpha, \beta x)$ is the lower **incomplete gamma function**.

If α is a positive **integer** (i.e., the distribution is an **Erlang distribution**), the cumulative distribution function has the following series expansion:^[5]

$$F(x; \alpha, \beta) = 1 - \sum_{i=0}^{\alpha-1} \frac{(\beta x)^i}{i!} e^{-\beta x} = e^{-\beta x} \sum_{i=\alpha}^{\infty} \frac{(\beta x)^i}{i!}$$

2.11.3 Properties

Skewness

The skewness is equal to $2/\sqrt{k}$, it depends only on the shape parameter (k) and approaches a normal distribution when k is large (approximately when $k > 10$).

Median calculation

Unlike the mode and the mean which have readily calculable formulas based on the parameters, the median does not have an easy closed form equation. The median for this distribution is defined as the value ν such that

$$\frac{1}{\Gamma(k)\theta^k} \int_0^\nu x^{k-1} e^{-\frac{x}{\theta}} dx = \frac{1}{2}.$$

A formula for approximating the median for any gamma distribution, when the mean is known, has been derived based on the fact that the ratio $\mu/(\mu - \nu)$ is approximately a linear function of k when $k \geq 1$.^[6] The approximation formula is

$$\nu \approx \mu \frac{3k - 0.8}{3k + 0.2},$$

where $\mu (= k\theta)$ is the mean.

A rigorous treatment of the problem of determining an asymptotic expansion and bounds for the median of the Gamma Distribution was handled first by Chen and Rubin, who proved

$$m - \frac{1}{3} < \lambda(m) < m,$$

where $\lambda(m)$ denotes the median of the $\text{Gamma}(m, 1)$ distribution.^[7]

K.P. Choi later showed that the first five terms in the asymptotic expansion of the median are

$$\lambda(m) = m - \frac{1}{3} + \frac{8}{405m} + \frac{184}{25515m^2} + \frac{2248}{3444525m^3} - \dots$$

by comparing the median to Ramanujan's θ function.^[8]

Later, it was shown that $\lambda(m)$ is a convex function of m .^[9]

Summation

If X_i has a $\text{Gamma}(k_i, \theta)$ distribution for $i = 1, 2, \dots, N$ (i.e., all distributions have the same scale parameter θ), then

$$\sum_{i=1}^N X_i \sim \text{Gamma}\left(\sum_{i=1}^N k_i, \theta\right)$$

provided all X_i are **independent**.

For the cases where the X_i are **independent** but have different scale parameters see Mathai (1982) and Moschopoulos (1984).

The gamma distribution exhibits **infinite divisibility**.

Scaling

If

$$X \sim \text{Gamma}(k, \theta),$$

then, for any $c > 0$,

$$cX \sim \text{Gamma}(k, c\theta).$$

Indeed, we know that if X is an **exponential r.v.** with rate λ then cX is an exponential r.v. with rate λ / c ; the same thing is valid with Gamma variates (and this can be checked using the **moment-generating function**, see, e.g., **these notes**, 10.4-(ii)): multiplication by a positive constant c divides the rate (or, equivalently, multiplies the scale).

Exponential family

The gamma distribution is a two-parameter **exponential family** with **natural parameters** $k - 1$ and $-1/\theta$ (equivalently, $\alpha - 1$ and $-\beta$), and **natural statistics** X and $\ln(X)$.

If the shape parameter k is held fixed, the resulting one-parameter family of distributions is a natural **exponential family**.

Logarithmic expectation

One can show that

$$\mathbf{E}[\ln(X)] = \psi(\alpha) - \ln(\beta)$$

or equivalently,

$$\mathbf{E}[\ln(X)] = \psi(k) + \ln(\theta)$$

where ψ is the digamma function.

This can be derived using the exponential family formula for the moment generating function of the sufficient statistic, because one of the sufficient statistics of the gamma distribution is $\ln(x)$.

Information entropy

The information entropy is

$$\mathbf{H}(X) = \mathbf{E}[-\ln(p(X))] = \mathbf{E}[-\alpha \ln(\beta) + \ln(\Gamma(\alpha)) - (\alpha - 1) \ln(X) + \beta X] = \alpha - \ln(\beta) + \ln(\Gamma(\alpha)) + (1 - \alpha)\psi(\alpha).$$

In the k, θ parameterization, the information entropy is given by

$$\mathbf{H}(X) = k + \ln(\theta) + \ln(\Gamma(k)) + (1 - k)\psi(k).$$

Kullback–Leibler divergence

The Kullback–Leibler divergence (KL-divergence), of $\text{Gamma}(\alpha_p, \beta_p)$ (“true” distribution) from $\text{Gamma}(\alpha_q, \beta_q)$ (“approximating” distribution) is given by^[10]

$$D_{\text{KL}}(\alpha_p, \beta_p; \alpha_q, \beta_q) = (\alpha_p - \alpha_q)\psi(\alpha_p) - \log \Gamma(\alpha_p) + \log \Gamma(\alpha_q) + \alpha_q(\log \beta_p - \log \beta_q) + \alpha_p \frac{\beta_q - \beta_p}{\beta_p}$$

Written using the k, θ parameterization, the KL-divergence of $\text{Gamma}(k_p, \theta_p)$ from $\text{Gamma}(k_q, \theta_q)$ is given by

$$D_{\text{KL}}(k_p, \theta_p; k_q, \theta_q) = (k_p - k_q)\psi(k_p) - \log \Gamma(k_p) + \log \Gamma(k_q) + k_q(\log \theta_q - \log \theta_p) + k_p \frac{\theta_p - \theta_q}{\theta_q}$$

Laplace transform

The Laplace transform of the gamma PDF is

$$F(s) = (1 + \theta s)^{-k} = \frac{\beta^\alpha}{(s + \beta)^\alpha}.$$

Differential equation

$$\left\{ \begin{aligned} \beta x f'(x) + f(x)(-\alpha\beta + \beta + x) &= 0; f(1) = \frac{e^{-1/\beta} \beta^{-\alpha}}{\Gamma(\alpha)} \\ x f'(x) + f(x)(-k + \theta x + 1) &= 0; f(1) = \frac{e^{-\theta} \left(\frac{1}{\theta}\right)^{-k}}{\Gamma(k)} \end{aligned} \right\}$$

2.11.4 Parameter estimation**Maximum likelihood estimation**

The likelihood function for N iid observations $(x_1, \dots, x_N)$ is

$$L(k, \theta) = \prod_{i=1}^N f(x_i; k, \theta)$$

from which we calculate the log-likelihood function

$$\ell(k, \theta) = (k-1) \sum_{i=1}^N \ln(x_i) - \sum_{i=1}^N \frac{x_i}{\theta} - Nk \ln(\theta) - N \ln(\Gamma(k))$$

Finding the maximum with respect to θ by taking the derivative and setting it equal to zero yields the **maximum likelihood estimator** of the θ parameter:

$$\hat{\theta} = \frac{1}{kN} \sum_{i=1}^N x_i$$

Substituting this into the log-likelihood function gives

$$\ell = (k-1) \sum_{i=1}^N \ln(x_i) - Nk - Nk \ln\left(\frac{\sum x_i}{kN}\right) - N \ln(\Gamma(k))$$

Finding the maximum with respect to k by taking the derivative and setting it equal to zero yields

$$\ln(k) - \psi(k) = \ln\left(\frac{1}{N} \sum_{i=1}^N x_i\right) - \frac{1}{N} \sum_{i=1}^N \ln(x_i)$$

There is no closed-form solution for k . The function is numerically very well behaved, so if a numerical solution is desired, it can be found using, for example, **Newton's method**. An initial value of k can be found either using the **method of moments**, or using the approximation

$$\ln(k) - \psi(k) \approx \frac{1}{2k} \left(1 + \frac{1}{6k+1}\right)$$

If we let

$$s = \ln \left(\frac{1}{N} \sum_{i=1}^N x_i \right) - \frac{1}{N} \sum_{i=1}^N \ln(x_i)$$

then k is approximately

$$k \approx \frac{3 - s + \sqrt{(s - 3)^2 + 24s}}{12s}$$

which is within 1.5% of the correct value.^[11] An explicit form for the Newton–Raphson update of this initial guess is:^[12]

$$k \leftarrow k - \frac{\ln(k) - \psi(k) - s}{\frac{1}{k} - \psi'(k)}.$$

Bayesian minimum mean squared error

With known k and unknown θ , the posterior density function for theta (using the standard scale-invariant prior for θ) is

$$P(\theta|k, x_1, \dots, x_N) \propto \frac{1}{\theta} \prod_{i=1}^N f(x_i; k, \theta)$$

Denoting

$$y \equiv \sum_{i=1}^N x_i, \quad P(\theta|k, x_1, \dots, x_N) = C(x_i) \theta^{-Nk-1} e^{-\frac{y}{\theta}}$$

Integration over θ can be carried out using a change of variables, revealing that $1/\theta$ is gamma-distributed with parameters $\alpha = Nk$, $\beta = y$.

$$\int_0^\infty \theta^{-Nk-1+m} e^{-\frac{y}{\theta}} d\theta = \int_0^\infty x^{Nk-1-m} e^{-xy} dx = y^{-(Nk-m)} \Gamma(Nk-m)$$

The moments can be computed by taking the ratio (m by $m = 0$)

$$\mathbf{E}[x^m] = \frac{\Gamma(Nk-m)}{\Gamma(Nk)} y^m$$

which shows that the mean $\pm$ standard deviation estimate of the posterior distribution for θ is

$$\frac{y}{Nk-1} \pm \frac{y^2}{(Nk-1)^2(Nk-2)}$$

2.11.5 Generating gamma-distributed random variables

Given the scaling property above, it is enough to generate gamma variables with $\theta = 1$ as we can later convert to any value of β with simple division.

Suppose we wish to generate random variables from $\text{Gamma}(n+\delta, 1)$, where n is a non-negative integer and $0 < \delta < 1$. Using the fact that a $\text{Gamma}(1, 1)$ distribution is the same as an $\text{Exp}(1)$ distribution, and noting the method of **generating exponential variables**, we conclude that if U is **uniformly distributed** on $(0, 1]$, then $-\ln(U)$ is distributed $\text{Gamma}(1, 1)$. Now, using the " α -addition" property of gamma distribution, we expand this result:

$$-\sum_{k=1}^n \ln U_k \sim \Gamma(n, 1)$$

where U_k are all uniformly distributed on $(0, 1]$ and **independent**. All that is left now is to generate a variable distributed as $\text{Gamma}(\delta, 1)$ for $0 < \delta < 1$ and apply the " α -addition" property once more. This is the most difficult part.

Random generation of gamma variates is discussed in detail by Devroye,^{[13]:401–428} noting that none are uniformly fast for all shape parameters. For small values of the shape parameter, the algorithms are often not valid.^{[13]:406} For arbitrary values of the shape parameter, one can apply the Ahrens and Dieter^[14] modified acceptance–rejection method Algorithm GD (shape $k \geq 1$), or transformation method^[15] when $0 < k < 1$. Also see Cheng and Feast Algorithm GKM 3^[16] or Marsaglia's squeeze method.^[17]

The following is a version of the Ahrens-Dieter **acceptance–rejection method**:^[14]

1. Let m be 1.
2. Generate V_{3m-2} , V_{3m-1} and V_{3m} as independent uniformly distributed on $(0, 1]$ variables.
3. If $V_{3m-2} \leq v_0$, where $v_0 = \frac{e}{e+\delta}$, then go to step 4, else go to step 5.
4. Let $\xi_m = V_{3m-1}^{1/\delta}$, $\eta_m = V_{3m} \xi_m^{\delta-1}$. Go to step 6.
5. Let $\xi_m = 1 - \ln V_{3m-1}$, $\eta_m = V_{3m} e^{-\xi_m}$.
6. If $\eta_m > \xi_m^{\delta-1} e^{-\xi_m}$, then increment m and go to step 2.
7. Assume $\xi = \xi_m$ to be the realization of $\Gamma(\delta, 1)$.

A summary of this is

$$\theta \left(\xi - \sum_{i=1}^{\lfloor k \rfloor} \ln(U_i) \right) \sim \Gamma(k, \theta)$$

where

- $\lfloor k \rfloor$ is the integral part of k ,
- ξ has been generated using the algorithm above with $\delta = \{k\}$ (the fractional part of k),
- U_k and V_l are distributed as explained above and are all independent.

While the above approach is technically correct, Devroye notes that it is linear in the value of k and in general is not a good choice. Instead he recommends using either rejection-based or table-based methods, depending on context.^{[13]:401–428}

For example, Marsaglia's simple transformation-rejection method relying on a one normal and one uniform random number:^[18]

1. Setup: $d=a-1/3$, $c=1/\text{sqrt}(9d)$.
2. Generate: $v=(1+c*x)^3$, with x standard normal.

3. if $v > 0$ and $\log(\text{UNI}) < 0.5 \cdot x^2 + d \cdot v + d \cdot \log(v)$ return $d \cdot v$.
4. go back to step 2.

With $1 \leq a = \alpha = k$ generates a gamma distributed random number in time that is approximately constant with k . The acceptance rate does depend on k , with an acceptance rate of 0.95, 0.98, and 0.99 for $k=1, 2$, and 4. For $k < 1$, one can use $\gamma_\alpha = \gamma_{1+\alpha} U^{1/\alpha}$ to boost k to be usable with this method.

2.11.6 Related distributions

Special cases

Conjugate prior

In Bayesian inference, the **gamma distribution** is the **conjugate prior** to many likelihood distributions: the **Poisson**, **exponential**, **normal** (with known mean), **Pareto**, gamma with known shape σ , **inverse gamma** with known shape parameter, and **Gompertz** with known scale parameter.

The gamma distribution's **conjugate prior** is:^[19]

$$p(k, \theta | p, q, r, s) = \frac{1}{Z} \frac{p^{k-1} e^{-\theta^{-1} q}}{\Gamma(k)^r \theta^{ks}},$$

where Z is the normalizing constant, which has no closed-form solution. The posterior distribution can be found by updating the parameters as follows:

$$\begin{aligned} p' &= p \prod_i x_i, \\ q' &= q + \sum_i x_i, \\ r' &= r + n, \\ s' &= s + n, \end{aligned}$$

where n is the number of observations, and x_i is the i th observation.

Compound gamma

If the shape parameter of the gamma distribution is known, but the inverse-scale parameter is unknown, then a gamma distribution for the inverse-scale forms a conjugate prior. The **compound distribution**, which results from integrating out the inverse-scale, has a closed form solution, known as the **compound gamma distribution**.^[20]

If instead the shape parameter is known but the mean is unknown, with the prior of the mean being given by another gamma distribution, then it results in **K-distribution**.

Others

- If $X \sim \text{Gamma}(1, 1/\lambda)$ (shape -scale parametrization), then X has an **exponential distribution** with rate parameter λ .
- If $X \sim \text{Gamma}(v/2, 2)$, then X is identical to $\chi^2(v)$, the **chi-squared distribution** with v degrees of freedom. Conversely, if $Q \sim \chi^2(v)$ and c is a positive constant, then $cQ \sim \text{Gamma}(v/2, 2c)$.
- If k is an integer, the gamma distribution is an **Erlang distribution** and is the probability distribution of the waiting time until the k th “arrival” in a one-dimensional Poisson process with intensity $1/\theta$. If

$$X \sim \Gamma(k \in \mathbf{Z}, \theta), \quad Y \sim \text{Pois}\left(\frac{x}{\theta}\right),$$

then

$$P(X > x) = P(Y < k).$$

- If X has a **Maxwell–Boltzmann distribution** with parameter a , then

$$X^2 \sim \Gamma\left(\frac{3}{2}, 2a^2\right)$$

- If $X \sim \text{Gamma}(k, \theta)$, then $\sqrt{X}$ follows a **generalized gamma distribution** with parameters $p=2$, $d=2k$, and $a = \sqrt{\theta}$.
- More generally, if $X \sim \text{Gamma}(k, \theta)$, then X^q for $q > 0$ follows a **generalized gamma distribution** with parameters $p = 1/q$, $d = k/q$, and $a = \theta^q$.
- If $X \sim \text{Gamma}(k, \theta)$, then $1/X \sim \text{Inv-Gamma}(k, \theta^{-1})$ (see **Inverse-gamma distribution** for derivation).
- If $X \sim \text{Gamma}(\alpha, \theta)$ and $Y \sim \text{Gamma}(\beta, \theta)$ are independently distributed, then $X/(X + Y)$ has a **beta distribution** with parameters α and β .
- If $X_i \sim \text{Gamma}(\alpha_i, 1)$ are independently distributed, then the vector $(X_1/S, \dots, X_n/S)$, where $S = X_1 + \dots + X_n$, follows a **Dirichlet distribution** with parameters $\alpha_1, \dots, \alpha_n$.
- For large k the gamma distribution converges to Gaussian distribution with mean $\mu = k\theta$ and variance $\sigma^2 = k\theta^2$.
- The gamma distribution is the **conjugate prior** for the precision of the **normal distribution** with known mean.
- The **Wishart distribution** is a multivariate generalization of the gamma distribution (samples are positive-definite matrices rather than positive real numbers).
- The gamma distribution is a special case of the **generalized gamma distribution**, the **generalized integer gamma distribution**, and the **generalized inverse Gaussian distribution**.
- Among the discrete distributions, the **negative binomial distribution** is sometimes considered the discrete analogue of the Gamma distribution.
- **Tweedie distributions** – the gamma distribution is a member of the family of Tweedie exponential dispersion models.

2.11.7 Applications

The gamma distribution has been used to model the size of insurance claims^[21] and rainfalls.^[22] This means that aggregate insurance claims and the amount of rainfall accumulated in a reservoir are modelled by a **gamma process**. The gamma distribution is also used to model errors in multi-level Poisson regression models, because the combination of the Poisson distribution and a gamma distribution is a **negative binomial distribution**.

In wireless communication, the gamma distribution is used to model the multi-path fading of signal power.

In neuroscience, the gamma distribution is often used to describe the distribution of inter-spike intervals.^{[23][24]}

In bacterial gene expression, the copy number of a constitutively expressed protein often follows the gamma distribution, where the scale and shape parameter are, respectively, the mean number of bursts per cell cycle and the mean number of protein molecules produced by a single mRNA during its lifetime.^[25]

In genomics, the gamma distribution was applied in peak calling step (i.e. in recognition of signal) in ChIP-chip^[26] and ChIP-seq^[27] data analysis.

The gamma distribution is widely used as a **conjugate prior** in Bayesian statistics. It is the conjugate prior for the precision (i.e. inverse of the variance) of a **normal distribution**. It is also the conjugate prior for the **exponential distribution**.

2.11.8 Notes

- [1] <http://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture6.pdf>
- [2] See Hogg and Craig (1978, Remark 3.3.1) for an explicit motivation
- [3] *Scalable Recommendation with Poisson Factorization*, Prem Gopalan, Jake M. Hofman, David Blei, arXiv.org 2014
- [4] Park, Sung Y.; Bera, Anil K. (2009). "Maximum entropy autoregressive conditional heteroskedasticity model" (PDF). *Journal of Econometrics* (Elsevier): 219–230. doi:10.1016/j.jeconom.2008.12.014. Retrieved 2011-06-02.
- [5] Papoulis, Pillai, *Probability, Random Variables, and Stochastic Processes*, Fourth Edition
- [6] Banneheka BMSG, Ekanayake GEMUPD (2009) "A new point estimator for the median of gamma distribution". *Viyodaya J Science*, 14:95–103
- [7] Jeesen Chen, Herman Rubin, Bounds for the difference between median and mean of gamma and poisson distributions, Statistics & Probability Letters, Volume 4, Issue 6, October 1986, Pages 281-283, ISSN 0167-7152, .
- [8] Choi, K.P. "On the Medians of the Gamma Distributions and an Equation of Ramanujan", Proceedings of the American Mathematical Society, Vol. 121, No. 1 (May, 1994), pp. 245-251.
- [9] Berg, Christian and Pedersen, Henrik L. "Convexity of the median in the gamma distribution".
- [10] W.D. Penny, [www.fil.ion.ucl.ac.uk/~{ }wpenny/publications/densities.ps KL-Divergences of Normal, Gamma, Dirichlet, and Wishart densities]
- [11] Minka, Thomas P. (2002) "Estimating a Gamma distribution". <http://research.microsoft.com/en-us/um/people/minka/papers/minka-gamma.pdf>
- [12] Choi, S.C.; Wette, R. (1969) "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias", *Technometrics*, 11(4) 683–690
- [13] Devroye, Luc (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag. ISBN 0-387-96305-7. See Chapter 9, Section 3.
- [14] Ahrens, J. H.; Dieter, U (January 1982). "Generating gamma variates by a modified rejection technique". *Communications of the ACM* **25** (1): 47–54. doi:10.1145/358315.358390.. See Algorithm GD, p. 53.
- [15] Ahrens, J. H.; Dieter, U. (1974). "Computer methods for sampling from gamma, beta, Poisson and binomial distributions". *Computing* **12**: 223–246. doi:10.1007/BF02293108. CiteSeerX: 10.1.1.93.3828.
- [16] Cheng, R.C.H., and Feast, G.M. Some simple gamma variate generators. *Appl. Stat.* 28 (1979), 290–295.
- [17] Marsaglia, G. The squeeze method for generating gamma variates. *Comput. Math. Appl.* 3 (1977), 321–325.
- [18] Marsaglia, G.; Tsang, W. W. (2000). "A simple method for generating gamma variables". *ACM Transactions on Mathematical Software* **26** (3): 363–372. doi:10.1145/358407.358414.
- [19] Fink, D. 1995 A Compendium of Conjugate Priors. In progress report: Extension and enhancement of methods for setting data quality objectives. (DOE contract 95-831).
- [20] Dubey, Satya D. (December 1970). "Compound gamma, beta and F distributions". *Metrika* **16**: 27–31. doi:10.1007/BF02613934.
- [21] p. 43, Philip J. Boland, Statistical and Probabilistic Methods in Actuarial Science, Chapman & Hall CRC 2007
- [22] Aksoy, H. (2000) "Use of Gamma Distribution in Hydrological Analysis", *Turk J. Engin Environ Sci*, 24, 419 – 428.
- [23] J. G. Robson and J. B. Troy, "Nature of the maintained discharge of Q, X, and Y retinal ganglion cells of the cat", *J. Opt. Soc. Am. A* 4, 2301–2307 (1987)
- [24] M.C.M. Wright, I.M. Winter, J.J. Forster, S. Bleeck "Response to best-frequency tone bursts in the ventral cochlear nucleus is governed by ordered inter-spike interval statistics", *Hearing Research* 317 (2014)
- [25] N. Friedman, L. Cai and X. S. Xie (2006) "Linking stochastic dynamics to population distribution: An analytical framework of gene expression", *Phys. Rev. Lett.* 97, 168302.

- [26] DJ Reiss, MT Facciotti and NS Baliga (2008) “Model-based deconvolution of genome-wide DNA binding”, *Bioinformatics*, 24, 396–403
- [27] MA Mendoza-Parra, M Nowicka, W Van Gool, H Gronemeyer (2013) “Characterising ChIP-seq binding patterns by model-based peak shape deconvolution”, *BMC Genomics*, 14:834

2.11.9 References

- R. V. Hogg and A. T. Craig (1978) *Introduction to Mathematical Statistics*, 4th edition. New York: Macmillan. (See Section 3.3.)
- P. G. Moschopoulos (1985) *The distribution of the sum of independent gamma random variables*, **Annals of the Institute of Statistical Mathematics**, 37, 541–544
- A. M. Mathai (1982) *Storage capacity of a dam with gamma type inputs*, **Annals of the Institute of Statistical Mathematics**, 34, 591–597

2.11.10 External links

- Hazewinkel, Michiel, ed. (2001), “Gamma-distribution”, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Weisstein, Eric W., “Gamma distribution”, *MathWorld*.
- Engineering Statistics Handbook

2.12 Generalized gamma distribution

The **generalized gamma distribution** is a continuous probability distribution with three parameters. It is a generalization of the two-parameter gamma distribution. Since many distributions commonly used for parametric models in survival analysis (such as the Weibull distribution and the log-normal distribution) are special cases of the generalized gamma, it is sometimes used to determine which parametric model is appropriate for a given set of data.^[1]

2.12.1 Characteristics

The generalized gamma has three parameters: $a > 0$, $d > 0$, and $p > 0$. For non-negative x , the probability density function of the generalized gamma is^[2]

$$f(x; a, d, p) = \frac{(p/a^d)x^{d-1}e^{-(x/a)^p}}{\Gamma(d/p)},$$

where $\Gamma(\cdot)$ denotes the gamma function.

The cumulative distribution function is

$$F(x; a, d, p) = \frac{\gamma(d/p, (x/a)^p)}{\Gamma(d/p)},$$

where $\gamma(\cdot)$ denotes the lower incomplete gamma function.

If $d = p$ then the generalized gamma distribution becomes the Weibull distribution. Alternatively, if $p = 1$ the generalised gamma becomes the gamma distribution.

Alternative parameterisations of this distribution are sometimes used; for example with the substitution $\alpha = d/p$.^[3] In addition, a shift parameter can be added, so the domain of x starts at some value other than zero.^[3] If the restrictions on the signs of a , d and p are also lifted (but $\alpha = d/p$ remains positive), this gives a distribution called the **Amoroso distribution**, after the Italian mathematician and economist Luigi Amoroso who described it in 1925.^[4]

2.12.2 Moments

If X has a generalized gamma distribution as above, then^[3]

$$E(X^r) = a^r \frac{\Gamma(\frac{d+r}{p})}{\Gamma(\frac{d}{p})}.$$

2.12.3 Kullback-Leibler divergence

If f_1 and f_2 are the probability density functions of two generalized gamma distributions, then their **Kullback-Leibler divergence** is given by

$$\begin{aligned} D_{KL}(f_1 \parallel f_2) &= \int_0^\infty f_1(x; a_1, d_1, p_1) \ln \frac{f_1(x; a_1, d_1, p_1)}{f_2(x; a_2, d_2, p_2)} dx \\ &= \ln \frac{p_1 a_2^{d_2} \Gamma(d_2/p_2)}{p_2 a_1^{d_1} \Gamma(d_1/p_1)} + \left[\frac{\psi(d_1/p_1)}{p_1} + \ln a_1 \right] (d_1 - d_2) + \frac{\Gamma((d_1 + p_2)/p_1)}{\Gamma(d_1/p_1)} \left(\frac{a_1}{a_2} \right)^{p_2} - \frac{d_1}{p_1} \end{aligned}$$

where $\psi(\cdot)$ is the digamma function.^[5]

2.12.4 Software implementation

In R implemented in the package *flexsurv*, function *dgengamma*, with different parametrisation: $\mu = \ln a + \frac{\ln d - \ln p}{p}$, $\sigma = \frac{1}{\sqrt{pd}}$, $Q = \sqrt{\frac{p}{d}}$.

2.12.5 See also

- Generalized integer gamma distribution

2.12.6 References

- [1] Box-Steffensmeier, Janet M.; Jones, Bradford S. (2004) *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press. ISBN 0-521-54673-7 (pp. 41-43)
- [2] Stacy, E.W. (1962). "A Generalization of the Gamma Distribution." *Annals of Mathematical Statistics* 33(3): 1187-1192. JSTOR 2237889
- [3] Johnson, N.L.; Kotz, S; Balakrishnan, N. (1994) *Continuous Univariate Distributions, Volume 1*, 2nd Edition. Wiley. ISBN 0-471-58495-9 (Section 17.8.7)
- [4] Gavin E. Crooks (2010), *The Amoroso Distribution*, Technical Note, Lawrence Berkeley National Laboratory.
- [5] C. Bauckhage (2014), Computing the Kullback-Leibler Divergence between two Generalized Gamma Distributions, arXiv: 1401.6853.

2.13 Generalized Pareto distribution

This article is about a particular family of continuous distributions referred to as the generalized Pareto distribution. For the hierarchy of generalized Pareto distributions, see [Pareto distribution](#).

In statistics, the **generalized Pareto distribution** (GPD) is a family of continuous probability distributions. It is often used to model the tails of another distribution. It is specified by three parameters: location μ , scale σ , and shape ξ .^{[1][2]} Sometimes it is specified by only scale and shape^[3] and sometimes only by its shape parameter. Some references give the shape parameter as $\kappa = -\xi$.^[4]

2.13.1 Definition

The standard cumulative distribution function (cdf) of the GPD is defined by^[5]

$$F_{\xi}(z) = \begin{cases} 1 - (1 + \xi z)^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - e^{-z} & \text{for } \xi = 0. \end{cases}$$

where the support is $z \geq 0$ for $\xi \geq 0$ and $0 \leq z \leq -1/\xi$ for $\xi < 0$.

$$f_{\xi}(z) = \begin{cases} (\xi z + 1)^{-\frac{\xi+1}{\xi}} & \text{for } \xi \neq 0, \\ e^{-z} & \text{for } \xi = 0. \end{cases}$$

Differential equation

The cdf of the GPD is a solution of the following differential equation:

$$\begin{cases} (\xi z + 1)f'_{\xi}(z) + (\xi + 1)f_{\xi}(z) = 0, \\ f_{\xi}(0) = 1 \end{cases}$$

2.13.2 Characterization

The related location-scale family of distributions is obtained by replacing the argument z by $\frac{x-\mu}{\sigma}$ and adjusting the support accordingly: The [cumulative distribution function](#) is

$$F_{(\xi, \mu, \sigma)}(x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{for } \xi = 0. \end{cases}$$

for $x \geq \mu$ when $\xi \geq 0$, and $\mu \leq x \leq \mu - \sigma/\xi$ when $\xi < 0$, where $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$.

The [probability density function](#) (pdf) is

$$f_{(\xi, \mu, \sigma)}(x) = \frac{1}{\sigma} \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)}$$

or equivalently

$$f_{(\xi, \mu, \sigma)}(x) = \frac{\sigma^{\frac{1}{\xi}}}{(\sigma + \xi(x - \mu))^{\frac{1}{\xi} + 1}}$$

again, for $x \geq \mu$ when $\xi \geq 0$, and $\mu \leq x \leq \mu - \sigma/\xi$ when $\xi < 0$.

The pdf is a solution of the following differential equation:

$$\left\{ \begin{array}{l} f'(x)(-\mu\xi + \sigma + \xi x) + (\xi + 1)f(x) = 0, \\ f(0) = \frac{(1 - \frac{\mu\xi}{\sigma})^{-\frac{1}{\xi} - 1}}{\sigma} \end{array} \right\}$$

2.13.3 Characteristic and Moment Generating Functions

The characteristic and moment generating functions are derived and skewness and kurtosis are obtained from MGF by Muraleedharan and Guedes Soares^[6]

2.13.4 Special cases

- If the shape ξ and location μ are both zero, the GPD is equivalent to the exponential distribution.
- With shape $\xi > 0$ and location $\mu = \sigma/\xi$, the GPD is equivalent to the Pareto distribution with scale $x_m = \sigma/\xi$ and shape $\alpha = 1/\xi$.

2.13.5 Generating generalized Pareto random variables

If U is uniformly distributed on $(0, 1]$, then

$$X = \mu + \frac{\sigma(U^{-\xi} - 1)}{\xi} \sim \text{GPD}(\mu, \sigma, \xi \neq 0)$$

and

$$X = \mu - \sigma \ln(U) \sim \text{GPD}(\mu, \sigma, \xi = 0).$$

Both formulas are obtained by inversion of the cdf.

In Matlab Statistics Toolbox, you can easily use “gprnd” command to generate generalized Pareto random numbers.

With GNU R you can use the packages POT or evd with the “rgpd” command (see for exact usage: <http://rss.acs.unt.edu/Rdoc/library/POT/html/simGPD.html>)

2.13.6 See also

- Pareto_distribution
- Generalized extreme value distribution
- Pickands–Balkema–de Haan theorem

2.13.7 Notes

- [1] Coles, Stuart (2001-12-12). *An Introduction to Statistical Modeling of Extreme Values*. Springer. p. 75. ISBN 9781852334598.
- [2] Dargahi-Noubary, G. R. (1989). “On tail estimation: An improved method”. *Mathematical Geology* **21** (8): 829–842. doi:10.1007/BF00894450.
- [3] Hosking, J. R. M.; Wallis, J. R. (1987). “Parameter and Quantile Estimation for the Generalized Pareto Distribution”. *Technometrics* **29** (3): 339–349. doi:10.2307/1269343.
- [4] Davison, A. C. (1984-09-30). “Modelling Excesses over High Thresholds, with an Application”. In de Oliveira, J. Tiago. *Statistical Extremes and Applications*. Kluwer. p. 462. ISBN 9789027718044.
- [5] Embrechts, Paul; Klüppelberg, Claudia; Mikosch, Thomas (1997-01-01). *Modelling extremal events for insurance and finance*. p. 162. ISBN 9783540609315.
- [6] Muraleedharan, G.; C, Guedes Soares (2014). “Characteristic and Moment Generating Functions of Generalised Pareto(GP3) and Weibull Distributions”. *Journal of Scientific Research and Reports* **3** (14): 1861–1874. doi:10.9734/JSRR/2014/10087.

2.13.8 References

- Pickands, James (1975). “Statistical inference using extreme order statistics”. *Annals of Statistics* **3**: 119–131. doi:10.1214/aos/1176343003
- Balkema, A.; De Haan, Laurens (1974). “Residual life time at great age”. *Annals of Probability* **2** (5): 792–804. doi:10.1214/aop/1176996548
- N. L. Johnson, S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions Volume 1, second edition*. New York: Wiley. ISBN 0-471-58495-9. Chapter 20, Section 12: Generalized Pareto Distributions.
- Barry C. Arnold (2011). “Chapter 7: Pareto and Generalized Pareto Distributions”. In Duangkamon Chotikapich. *Modeling Distributions and Lorenz Curves*. New York: Springer. ISBN 9780387727967.
- Arnold, B. C. and Laguna, L. (1977). *On generalized Pareto distributions with applications to income data*. Ames, Iowa: Iowa State University, Department of Economics.

2.13.9 External links

- Mathworks: Generalized Pareto distribution

2.14 Gamma/Gompertz distribution

In probability and statistics, the **Gamma/Gompertz distribution** is a continuous probability distribution. It has been used as an aggregate-level model of customer lifetime and a model of mortality risks.

2.14.1 Specification

Probability density function

The probability density function of the Gamma/Gompertz distribution is:

$$f(x; b, s, \beta) = \frac{b s e^{b x} \beta^s}{(\beta - 1 + e^{b x})^{s+1}}$$

where $b > 0$ is the scale parameter and $\beta, s > 0$ are the shape parameters of the Gamma/Gompertz distribution.

Cumulative distribution function

The **cumulative distribution function** of the Gamma/Gompertz distribution is:

$$\begin{aligned} F(x; b, s, \beta) &= 1 - \frac{\beta^s}{(\beta - 1 + e^{bx})^s}, \quad x > 0, \quad b, s, \beta > 0 \\ &= 1 - e^{-bsx}, \quad \beta = 1 \end{aligned}$$

Moment generating function

The moment generating function is given by:

$$E(e^{-tx}) = \begin{cases} \beta^s \frac{sb}{t + sb} {}_2F_1(s + 1, (t/b) + s; (t/b) + s + 1; 1 - \beta), & \beta \neq 1; \\ \frac{sb}{t + sb}, & \beta = 1. \end{cases}$$

where ${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} [(a)_k (b)_k / (c)_k] z^k / k!$ is a **Hypergeometric function**.

2.14.2 Properties

The Gamma/Gompertz distribution is a flexible distribution that can be skewed to the right or to the left.

2.14.3 Related distributions

- When $\beta = 1$, this reduces to an **Exponential distribution** with parameter sb .
- The **gamma distribution** is a natural **conjugate prior** to a Gompertz likelihood with known, scale parameter b . ^[1]
- When the shape parameter η of a **Gompertz distribution** varies according to a **gamma distribution** with shape parameter α and scale parameter β (mean = α/β), the distribution of x is Gamma/Gompertz. ^[1]

2.14.4 See also

- **Gompertz distribution**
- **Customer lifetime value**

2.14.5 Notes

[1] Bemmaor, A.C.; Glady, N. (2012)

2.14.6 References

- Bemmaor, Albert C.; Glady, Nicolas (2012). “Modeling Purchasing Behavior With Sudden 'Death': A Flexible Customer Lifetime Model”. *Management Science* **58** (5): 1012–1021. doi:10.1287/mnsc.1110.1461.
- Bemmaor, Albert C.; Glady, Nicolas (2011). “Implementing the Gamma/Gompertz/NBD Model in MATLAB” (PDF). Cergy-Pontoise: ESSEC Business School.

- Gompertz, B. (1825). “On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies”. *Philosophical Transactions of the Royal Society of London* **115**: 513–583. doi:10.1098/rstl.1825.0026. JSTOR 107756.
- Johnson, Norman L.; Kotz, Samuel; Balakrishnan, N. (1995). “Continuous Univariate Distributions” **2** (2nd ed.). New York: John Wiley & Sons. pp. 25–26. ISBN 0-471-58494-0.
- Manton, K. G.; Stallard, E.; Vaupel, J. W. (1986). “Alternative Models for the Heterogeneity of Mortality Risks Among the Aged”. *Journal of the American Statistical Association* **81**: 635–644. doi:10.1080/01621459.1986.10478316.

2.15 Gompertz distribution

In probability and statistics, the **Gompertz distribution** is a continuous probability distribution. The Gompertz distribution is often applied to describe the distribution of adult lifespans by demographers^{[1][2]} and actuaries.^{[3][4]} Related fields of science such as biology^[5] and gerontology^[6] also considered the Gompertz distribution for the analysis of survival. More recently, computer scientists have also started to model the failure rates of computer codes by the Gompertz distribution.^[7] In Marketing Science, it has been used as an individual-level simulation for customer lifetime value modeling.^[8]

2.15.1 Specification

Probability density function

The probability density function of the Gompertz distribution is:

$$f(x; \eta, b) = b\eta e^{bx} e^{\eta} \exp(-\eta e^{bx}) \text{ for } x \geq 0,$$

where $b > 0$ is the scale parameter and $\eta > 0$ is the shape parameter of the Gompertz distribution. In the actuarial and biological sciences and in demography, the Gompertz distribution is parametrized slightly differently (Gompertz–Makeham law of mortality).

Cumulative distribution function

The cumulative distribution function of the Gompertz distribution is:

$$F(x; \eta, b) = 1 - \exp(-\eta(e^{bx} - 1)),$$

where $\eta, b > 0$, and $x \geq 0$.

Moment generating function

The moment generating function is:

$$E(e^{-tX}) = \eta e^{\eta} E_{t/b}(\eta)$$

where

$$E_{t/b}(\eta) = \int_1^{\infty} e^{-\eta v} v^{-t/b} dv, \quad t > 0.$$

2.15.2 Properties

The Gompertz distribution is a flexible distribution that can be skewed to the right and to the left.

Shapes

The Gompertz density function can take on different shapes depending on the values of the shape parameter η :

- When $\eta \geq 1$, the probability density function has its mode at 0.
- When $0 < \eta < 1$, the probability density function has its mode at

$$x^* = (1/b) \ln(1/\eta) \text{ with } 0 < F(x^*) < 1 - e^{-1} = 0.632121$$

Kullback-Leibler divergence

If f_1 and f_2 are the probability density functions of two Gompertz distributions, then their **Kullback-Leibler divergence** is given by

$$\begin{aligned} D_{KL}(f_1 \parallel f_2) &= \int_0^\infty f_1(x; b_1, \eta_1) \ln \frac{f_1(x; b_1, \eta_1)}{f_2(x; b_2, \eta_2)} dx \\ &= \ln \frac{e^{\eta_1} b_1 \eta_1}{e^{\eta_2} b_2 \eta_2} + e^{\eta_1} \left[\left(\frac{b_2}{b_1} - 1 \right) \text{Ei}(-\eta_1) + \frac{\eta_2}{\eta_1^{\frac{b_2}{b_1}}} \Gamma\left(\frac{b_2}{b_1} + 1, \eta_1\right) \right] - (\eta_1 + 1) \end{aligned}$$

where $\text{Ei}(\cdot)$ denotes the **exponential integral** and $\Gamma(\cdot, \cdot)$ is the **upper incomplete gamma function**.^[9]

2.15.3 Related distributions

- If X is defined to be the result of sampling from a **Gumbel distribution** until a negative value Y is produced, and setting $X = -Y$, then X has a Gompertz distribution.
- The **gamma distribution** is a natural **conjugate prior** to a Gompertz likelihood with known scale parameter b .^[8]
- When η varies according to a **gamma distribution** with shape parameter α and scale parameter β (mean = α/β), the distribution of x is Gamma/Gompertz.^[8]

2.15.4 See also

- Gompertz function
- Customer lifetime value
- Gamma Gompertz distribution

2.15.5 Notes

- [1] Vaupel, James W. (1986). "How change in age-specific mortality affects life expectancy". *Population Studies* **40** (1): 147–157. doi:10.1080/0032472031000141896.
- [2] Preston, Samuel H.; Heuveline, Patrick; Guillot, Michel (2001). *Demography: measuring and modeling population processes*. Oxford: Blackwell.
- [3] Benjamin, Bernard; Haycocks, H.W.; Pollard, J. (1980). *The Analysis of Mortality and Other Actuarial Statistics*. London: Heinemann.
- [4] Willemse, W. J.; Koppelaar, H. (2000). "Knowledge elicitation of Gompertz' law of mortality". *Scandinavian Actuarial Journal* (2): 168–179.
- [5] Economos, A. (1982). "Rate of aging, rate of dying and the mechanism of mortality". *Archives of Gerontology and Geriatrics* **1** (1): 46–51.
- [6] Brown, K.; Forbes, W. (1974). "A mathematical model of aging processes". *Journal of Gerontology* **29** (1): 46–51. doi:10.1093/geronj/29.1.46.
- [7] Ohishi, K.; Okamura, H.; Dohi, T. (2009). "Gompertz software reliability model: estimation algorithm and empirical validation". *Journal of Systems and Software* **82** (3): 535–543. doi:10.1016/j.jss.2008.11.840.
- [8] Bemmaor, Albert C.; Glady, Nicolas (2012). "Modeling Purchasing Behavior With Sudden 'Death': A Flexible Customer Lifetime Model". *Management Science* **58** (5): 1012–1021. doi:10.1287/mnsc.1110.1461.
- [9] Bauckhage, C. (2014), Characterizations and Kullback-Leibler Divergence of Gompertz Distributions, arXiv:1402.3193.

2.15.6 References

- Bemmaor, Albert C.; Glady, Nicolas (2011). "Implementing the Gamma/Gompertz/NBD Model in MATLAB" (PDF). Cergy-Pontoise: ESSEC Business School.
- Gompertz, B. (1825). "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies". *Philosophical Transactions of the Royal Society of London* **115**: 513–583. doi:10.1098/rstl.1825.0026. JSTOR 107756.
- Johnson, Norman L.; Kotz, Samuel; Balakrishnan, N. (1995). "Continuous Univariate Distributions" **2** (2nd ed.). New York: John Wiley & Sons. pp. 25–26. ISBN 0-471-58494-0.
- Sheikh, A. K.; Boah, J. K.; Younas, M. (1989). "Truncated Extreme Value Model for Pipeline Reliability". *Reliability Engineering and System Safety* **25** (1): 1–14. doi:10.1016/0951-8320(89)90020-3.

2.16 Half-normal distribution

The **half-normal distribution** is a special case of the **folded normal distribution**.

Let X follow an ordinary **normal distribution**, $N(0, \sigma^2)$, then $Y = |X|$ follows a half-normal distribution. Thus, the half-normal distribution is a fold at the mean of an ordinary normal distribution with mean zero.

Using the σ parametrization of the normal distribution, the **probability density function** (PDF) of the half-normal is given by

$$f_Y(y; \sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right) \quad y > 0$$

Where $E[Y] = \mu = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$.

Alternatively using a scaled precision (inverse of the variance) parametrization (to avoid issues if σ is near zero), obtained by setting $\theta = \frac{\sqrt{\pi}}{\sigma\sqrt{2}}$, the probability density function is given by

$$f_Y(y; \theta) = \frac{2\theta}{\pi} \exp\left(-\frac{y^2\theta^2}{\pi}\right) \quad y > 0$$

where $E[Y] = \mu = \frac{1}{\theta}$.

The cumulative distribution function (CDF) is given by

$$F_Y(y; \sigma) = \int_0^y \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

Using the change-of-variables $z = x/(\sqrt{2}\sigma)$, the CDF can be written as

$$F_Y(y; \sigma) = \frac{2}{\sqrt{\pi}} \int_0^{y/(\sqrt{2}\sigma)} \exp(-z^2) dz = \operatorname{erf}\left(\frac{y}{\sqrt{2}\sigma}\right),$$

where $\operatorname{erf}(x)$ is the error function, a standard function in many mathematical software packages.

The quantile function (or inverse CDF) is written:

$$Q(F; \sigma) = \sigma\sqrt{2} \operatorname{erf}^{-1}(F)$$

where $0 \leq F \leq 1$ and $\operatorname{erf}^{-1}()$ is the inverse error function

The expectation is then given by

$$E(Y) = \sigma\sqrt{2/\pi},$$

The variance is given by

$$\operatorname{Var}(Y) = \sigma^2 \left(1 - \frac{2}{\pi}\right).$$

Since this is proportional to the variance σ^2 of X , σ can be seen as a scale parameter of the new distribution.

The entropy of the half-normal distribution is exactly one bit less the entropy of a zero-mean normal distribution with the same second moment about 0. This can be understood intuitively since the magnitude operator reduces information by one bit (if the probability distribution at its input is even). Alternatively, since a half-normal distribution is always positive, the one bit it would take to record whether a standard normal random variable were positive (say, a 1) or negative (say, a 0) is no longer necessary. Thus,

$$H(Y) = \frac{1}{2} \log\left(\frac{\pi\sigma^2}{2}\right) + \frac{1}{2}$$

Differential equation

$$\left\{ \begin{aligned} \sigma^2 f'(x) + x f(x) &= 0, f(1) = \frac{\sqrt{\frac{2}{\pi}} e^{-\frac{1}{2\sigma^2}}}{\sigma} \\ \pi f'(x) + 2\theta^2 x f(x) &= 0, f(1) = \frac{2e^{-\frac{\theta^2}{\pi}} \theta}{\pi} \end{aligned} \right\}$$

2.16.1 Parameter estimation

Given numbers $\{x_i\}_{i=1}^n$ drawn from a half-normal distribution, the unknown parameter σ of that distribution can be estimated by the method of **maximum likelihood**, giving

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

2.16.2 Related distributions

- The distribution is a special case of the **folded normal distribution** with $\mu = 0$.
- It also coincides with a zero-mean normal distribution truncated from below at zero (see **truncated normal distribution**)
- $(Y/\sigma)^2$ has a **chi square distribution** with 1 degree of freedom.
- Y/σ has a **chi distribution** with 1 degree of freedom, if $Y \sim HN(\sigma)$ then $Y/\sigma \sim \chi_1$.

2.16.3 External links

- **Half-Normal Distribution at MathWorld**

(note that MathWorld uses the parameter $\theta = \frac{1}{\sigma} \sqrt{\pi/2}$)

2.16.4 References

2.17 Hotelling's T-squared distribution

In statistics **Hotelling's T-squared distribution** is a univariate distribution proportional to the **F-distribution** and arises importantly as the **distribution** of a set of statistics which are natural generalizations of the statistics underlying Student's *t*-distribution. In particular, the distribution arises in **multivariate statistics** in undertaking **tests** of the differences between the (multivariate) means of different populations, where tests for univariate problems would make use of a *t*-test.

The distribution is named for **Harold Hotelling**, who developed it^[1] as a generalization of Student's *t*-distribution.

2.17.1 The distribution

If the vector $p\mathbf{d}$ is **Gaussian multivariate-distributed** with zero mean and unit **covariance matrix** $\mathbf{N}(p\mathbf{0}, p\mathbf{I}_p)$ and $m\mathbf{M}p$ is a $p \times p$ matrix with a **Wishart distribution** with unit **scale matrix** and m **degrees of freedom** $\mathbf{W}(p\mathbf{I}_p, m)$ then $m(\mathbf{d}'p\mathbf{M}^{-1}p\mathbf{d})$ has a Hotelling T^2 distribution with dimensionality parameter p and m degrees of freedom.^[2]

If the notation $T_{p,m}^2$ is used to denote a random variable having Hotelling's *T*-squared distribution with parameters p and m then, if a random variable X has Hotelling's *T*-squared distribution,

$$X \sim T_{p,m}^2$$

then^[1]

$$\frac{m-p+1}{pm} X \sim F_{p, m-p+1}$$

where $F_{p, m-p+1}$ is the F -distribution with parameters p and $m-p+1$.

2.17.2 Hotelling's T -squared statistic

Hotelling's T -squared statistic is a generalization of Student's t statistic that is used in multivariate hypothesis testing, and is defined as follows.^[1]

Let $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a p -variate normal distribution with location $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Let

$$\mathbf{x}_1, \dots, \mathbf{x}_n \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

be n independent random variables, which may be represented as $p \times 1$ column vectors of real numbers. Define

$$\bar{\mathbf{x}} = \frac{\mathbf{x}_1 + \dots + \mathbf{x}_n}{n}$$

to be the sample mean. It can be shown that

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim \chi_p^2,$$

where χ_p^2 is the chi-squared distribution with p degrees of freedom. To show this use the fact that $\bar{\mathbf{x}} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$ and then derive the characteristic function of the random variable $\mathbf{y} = n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$. This is done below,

$$\begin{aligned} \phi_{\mathbf{y}}(\theta) &= \mathbb{E} e^{i\theta \mathbf{y}}, \\ &= \mathbb{E} e^{i\theta n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})} \\ &= \int e^{i\theta n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}/n|^{-\frac{1}{2}} e^{-\frac{1}{2} n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})} dx_1 \dots dx_p \\ &= \int (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}/n|^{-\frac{1}{2}} e^{-\frac{1}{2} n(\bar{\mathbf{x}} - \boldsymbol{\mu})' (\boldsymbol{\Sigma}^{-1} - 2i\theta \boldsymbol{\Sigma}^{-1}) (\bar{\mathbf{x}} - \boldsymbol{\mu})} dx_1 \dots dx_p, \\ &= |(\boldsymbol{\Sigma}^{-1} - 2i\theta \boldsymbol{\Sigma}^{-1})^{-1}/n|^{\frac{1}{2}} |\boldsymbol{\Sigma}/n|^{-\frac{1}{2}} \int (2\pi)^{-\frac{p}{2}} |(\boldsymbol{\Sigma}^{-1} - 2i\theta \boldsymbol{\Sigma}^{-1})^{-1}/n|^{-\frac{1}{2}} e^{-\frac{1}{2} n(\bar{\mathbf{x}} - \boldsymbol{\mu})' (\boldsymbol{\Sigma}^{-1} - 2i\theta \boldsymbol{\Sigma}^{-1}) (\bar{\mathbf{x}} - \boldsymbol{\mu})} dx_1 \dots dx_p, \\ &= |(\mathbf{I}_p - 2i\theta \mathbf{I}_p)|^{-\frac{1}{2}}, \\ &= (1 - 2i\theta)^{-\frac{p}{2}}. \blacksquare \end{aligned}$$

However, $\boldsymbol{\Sigma}$ is often unknown and we wish to do hypothesis testing on the location $\boldsymbol{\mu}$.

Sum of p squared t's

Define

$$\mathbf{W} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

to be the **sample covariance**. Here we denote **transpose** by an apostrophe. It can be shown that $\mathbf{W}$ is a **positive (semi) definite** matrix and $(n-1)\mathbf{W}$ follows a p -variate **Wishart distribution** with $n-1$ degrees of freedom.^[3] Hotelling's T -squared statistic is then defined^[4] to be

$$t^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{W}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

and, also from **above**,

$$t^2 \sim T_{p,n-1}^2$$

i.e.

$$\frac{n-p}{p(n-1)} t^2 \sim F_{p,n-p},$$

where $F_{p,n-p}$ is the **F-distribution** with parameters p and $n-p$. In order to calculate a p value, multiply the t^2 statistic by the above constant and use the F -distribution.

2.17.3 Hotelling's two-sample T -squared statistic

If $\mathbf{x}_1, \dots, \mathbf{x}_{n_x} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ and $\mathbf{y}_1, \dots, \mathbf{y}_{n_y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$, with the samples **independently** drawn from two **independent multivariate normal distributions** with the same mean and covariance, and we define

$$\bar{\mathbf{x}} = \frac{1}{n_x} \sum_{i=1}^{n_x} \mathbf{x}_i \quad \bar{\mathbf{y}} = \frac{1}{n_y} \sum_{i=1}^{n_y} \mathbf{y}_i$$

as the sample means, and

$$\mathbf{W} = \frac{\sum_{i=1}^{n_x} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' + \sum_{i=1}^{n_y} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'}{n_x + n_y - 2}$$

as the unbiased pooled covariance matrix estimate, then **Hotelling's two-sample T -squared statistic** is

$$t^2 = \frac{n_x n_y}{n_x + n_y} (\bar{\mathbf{x}} - \bar{\mathbf{y}})' \mathbf{W}^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}}) \sim T^2(p, n_x + n_y - 2)$$

and it can be related to the F -distribution by^[3]

$$\frac{n_x + n_y - p - 1}{(n_x + n_y - 2)p} t^2 \sim F(p, n_x + n_y - 1 - p).$$

The non-null distribution of this statistic is the **noncentral F-distribution** (the ratio of a **non-central Chi-squared** random variable and an independent central **Chi-squared** random variable)

$$\frac{n_x + n_y - p - 1}{(n_x + n_y - 2)p} t^2 \sim F(p, n_x + n_y - 1 - p; \delta),$$

with

$$\delta = \frac{n_x n_y}{n_x + n_y} \boldsymbol{\nu}' \mathbf{V}^{-1} \boldsymbol{\nu},$$

where $\boldsymbol{\nu}$ is the difference vector between the population means.

More robust and powerful tests than Hotelling's two-sample test have been proposed in the literature, see for example the interpoint distance based tests which can be applied also when the number of variables is comparable with, or even larger than, the number of subjects.^{[5][6]}

In the two variable case, the formula simplifies nicely allowing appreciation of how the correlation, r , between the variables affects t^2 . If we define

$$d_1 = \bar{x}_{.1} - \bar{y}_{.1}, \quad d_2 = \bar{x}_{.2} - \bar{y}_{.2}$$

and

$$SD_1 = \sqrt{W_{11}} \quad SD_2 = \sqrt{W_{22}}$$

then

$$t^2 = \frac{n_x n_y}{(n_x + n_y)(1 - r^2)} \left[\left(\frac{d_1}{SD_1} \right)^2 + \left(\frac{d_2}{SD_2} \right)^2 - 2r \left(\frac{d_1}{SD_1} \right) \left(\frac{d_2}{SD_2} \right) \right]$$

Thus, if the differences in the two rows of the vector $(\bar{\mathbf{x}} - \bar{\mathbf{y}})$ are of the same sign, in general, t^2 becomes smaller as r becomes more positive. If the differences are of opposite sign t^2 becomes larger as r becomes more positive.

2.17.4 See also

- **Student's t -test** in univariate statistics
- **Student's t -distribution** in univariate probability theory
- **Multivariate Student distribution.**
- **F -distribution** (commonly tabulated or available in software libraries, and hence used for testing the T -squared statistic using the relationship given above)
- **Wilks' lambda distribution** (in multivariate statistics Wilks's Λ is to Hotelling's T^2 as Snedecor's F is to Student's t in univariate statistics).

2.17.5 References

- [1] Hotelling, H. (1931). “The generalization of Student’s ratio”. *Annals of Mathematical Statistics* **2** (3): 360–378. doi:10.1214/aoms/1177732979.
- [2] Eric W. Weisstein, *CRC Concise Encyclopedia of Mathematics, Second Edition*, Chapman & Hall/CRC, 2003, p. 1408
- [3] Mardia, K. V.; Kent, J. T.; Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press. ISBN 0-12-471250-9.
- [4] <http://www.itl.nist.gov/div898/handbook/pmc/section5/pmc543.htm>
- [5] Marozzi, M. (2014). “Multivariate tests based on interpoint distances with application to magnetic resonance imaging”. *Statistical Methods in Medical Research*. doi:10.1177/0962280214529104.
- [6] Marozzi, M. (2015). “Multivariate multidistance tests for high-dimensional low sample size case-control studies”. *Statistics in Medicine* **34**. doi:10.1002/sim.6418.

2.17.6 External links

- Prokhorov, A. V. (2001), “Hotelling T^2 -distribution”, in Hazewinkel, Michiel, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4

2.18 Inverse Gaussian distribution

In probability theory, the **inverse Gaussian distribution** (also known as the **Wald distribution**) is a two-parameter family of continuous probability distributions with support on $(0, \infty)$.

Its probability density function is given by

$$f(x; \mu, \lambda) = \left[\frac{\lambda}{2\pi x^3} \right]^{1/2} \exp \frac{-\lambda(x - \mu)^2}{2\mu^2 x}$$

for $x > 0$, where $\mu > 0$ is the mean and $\lambda > 0$ is the shape parameter.

As λ tends to infinity, the inverse Gaussian distribution becomes more like a normal (Gaussian) distribution. The inverse Gaussian distribution has several properties analogous to a Gaussian distribution. The name can be misleading: it is an “inverse” only in that, while the Gaussian describes a Brownian Motion’s level at a fixed time, the inverse Gaussian describes the distribution of the time a Brownian Motion with positive drift takes to reach a fixed positive level.

Its cumulant generating function (logarithm of the characteristic function) is the inverse of the cumulant generating function of a Gaussian random variable.

To indicate that a random variable X is inverse Gaussian-distributed with mean μ and shape parameter λ we write

$$X \sim IG(\mu, \lambda).$$

2.18.1 Properties

Summation

If X_i has a $IG(\mu_0 w_i, \lambda_0 w_i^2)$ distribution for $i = 1, 2, \dots, n$ and all X_i are independent, then

$$S = \sum_{i=1}^n X_i \sim IG \left(\mu_0 \sum w_i, \lambda_0 \left(\sum w_i \right)^2 \right).$$

Note that

$$\frac{\text{Var}(X_i)}{\text{E}(X_i)} = \frac{\mu_0^2 w_i^2}{\lambda_0 w_i^2} = \frac{\mu_0^2}{\lambda_0}$$

is constant for all i . This is a **necessary condition** for the summation. Otherwise S would not be inverse Gaussian.

Scaling

For any $t > 0$ it holds that

$$X \sim IG(\mu, \lambda) \Rightarrow tX \sim IG(t\mu, t\lambda).$$

Exponential family

The inverse Gaussian distribution is a two-parameter **exponential family** with natural parameters $-\lambda/(2\mu^2)$ and $-\lambda/2$, and natural statistics X and $1/X$.

Differential equation

Main article: **Differential equation**

$$\left\{ 2\mu^2 x^2 f'(x) + f(x) (-\lambda\mu^2 + \lambda x^2 + 3\mu^2 x) = 0, f(1) = \frac{\sqrt{\lambda} e^{-\frac{\lambda(1-\mu)^2}{2\mu^2}}}{\sqrt{2\pi}} \right\}$$

2.18.2 Relationship with Brownian motion

The **stochastic process** X_t given by

$$X_0 = 0$$

$$X_t = \nu t + \sigma W_t$$

(where W_t is a standard **Brownian motion** and $\nu > 0$) is a Brownian motion with drift ν .

Then, the **first passage time** for a fixed level $\alpha > 0$ by X_t is distributed according to an inverse-gaussian:

$$T_\alpha = \inf\{0 < t \mid X_t = \alpha\} \sim IG\left(\frac{\alpha}{\nu}, \frac{\alpha^2}{\sigma^2}\right).$$

When drift is zero

A common special case of the above arises when the Brownian motion has no drift. In that case, parameter μ tends to infinity, and the first passage time for fixed level α has probability density function

$$f\left(x; 0, \left(\frac{\alpha}{\sigma}\right)^2\right) = \frac{\alpha}{\sigma\sqrt{2\pi x^3}} \exp\left(-\frac{\alpha^2}{2x\sigma^2}\right).$$

This is a **Lévy distribution** with parameters $c = \frac{\alpha^2}{\sigma^2}$ and $\mu = 0$.

2.18.3 Maximum likelihood

The model where

$$X_i \sim IG(\mu, \lambda w_i), \quad i = 1, 2, \dots, n$$

with all w_i known, (μ, λ) unknown and all X_i independent has the following likelihood function

$$L(\mu, \lambda) = \left(\frac{\lambda}{2\pi} \right)^{\frac{n}{2}} \left(\prod_{i=1}^n \frac{w_i}{X_i^3} \right)^{\frac{1}{2}} \exp \left(\frac{\lambda}{\mu} \sum_{i=1}^n w_i - \frac{\lambda}{2\mu^2} \sum_{i=1}^n w_i X_i - \frac{\lambda}{2} \sum_{i=1}^n w_i \frac{1}{X_i} \right).$$

Solving the likelihood equation yields the following maximum likelihood estimates

$$\hat{\mu} = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i}, \quad \frac{1}{\hat{\lambda}} = \frac{1}{n} \sum_{i=1}^n w_i \left(\frac{1}{X_i} - \frac{1}{\hat{\mu}} \right).$$

$\hat{\mu}$ and $\hat{\lambda}$ are independent and

$$\hat{\mu} \sim IG \left(\mu, \lambda \sum_{i=1}^n w_i \right) \quad \frac{n}{\hat{\lambda}} \sim \frac{1}{\lambda} \chi_{n-1}^2.$$

2.18.4 Generating random variates from an inverse-Gaussian distribution

The following algorithm may be used.^[1]

Generate a random variate from a normal distribution with a mean of 0 and 1 standard deviation

$$\nu = N(0, 1).$$

Square the value

$$y = \nu^2$$

and use this relation

$$x = \mu + \frac{\mu^2 y}{2\lambda} - \frac{\mu}{2\lambda} \sqrt{4\mu\lambda y + \mu^2 y^2}.$$

Generate another random variate, this time sampled from a uniform distribution between 0 and 1

$$z = U(0, 1).$$

If

$$z \leq \frac{\mu}{\mu + x}$$

then return

x

else return

$$\frac{\mu^2}{x}.$$

Sample code in **Java**:

```
1 public double inverseGaussian(double mu, double lambda) {
2   Random rand = new Random();
3   double v = rand.nextGaussian();
4   double y = v*v;
5   double x = mu + (mu*mu*y)/(2*lambda) - (mu/(2*lambda)) * Math.sqrt(4*mu*lambda*y + mu*mu*y*y);
6   double test = rand.nextDouble();
7   // sample from a uniform distribution between 0 and 1
8   if (test <= (mu)/(mu + x)) return x;
9   else return (mu*mu)/x;
10 }
11 }
```

And to plot Wald distribution in **Python** using **matplotlib** and **NumPy**:

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 4 h = plt.hist(np.random.wald(3, 2, 100000), bins=200,
4 normed=True)
5 6 plt.show()
```

2.18.5 Related distributions

- If $X \sim \text{IG}(\mu, \lambda)$ then $kX \sim \text{IG}(k\mu, k\lambda)$
- If $X_i \sim \text{IG}(\mu, \lambda)$ then $\sum_{i=1}^n X_i \sim \text{IG}(n\mu, n^2\lambda)$
- If $X_i \sim \text{IG}(\mu, \lambda)$ for $i = 1, \dots, n$ then $\bar{X} \sim \text{IG}(\mu, n\lambda)$
- If $X_i \sim \text{IG}(\mu_i, 2\mu_i^2)$ then $\sum_{i=1}^n X_i \sim \text{IG}\left(\sum_{i=1}^n \mu_i, 2\left(\sum_{i=1}^n \mu_i\right)^2\right)$

The convolution of a inverse Gaussian distribution (a Wald distribution) and an exponential (an ex-Wald distribution) is used as a model for response times in psychology,^[2] with visual search as one example.^[3]

2.18.6 History

This distribution appears to have been first derived by Schrödinger in 1915 as the time to first passage of a Brownian motion.^[4] The name inverse Gaussian was proposed by Tweedie in 1945.^[5] Wald re-derived this distribution in 1947 as the limiting form of a sample in a sequential probability ratio test. Tweedie investigated this distribution in 1957 and established some of its statistical properties.

2.18.7 Software

The R programming language has software for this distribution.^[6] The inverse Gaussian distribution can be called using the `statmod` package.

2.18.8 See also

- Generalized inverse Gaussian distribution
- Tweedie distributions—The inverse Gaussian distribution is a member of the family of Tweedie exponential dispersion models
- Stopping time

2.18.9 Notes

- [1] Michael, John R.; Schucany, William R.; Haas, Roy W. (May 1976). "Generating Random Variates Using Transformations with Multiple Roots". *The American Statistician* (American Statistical Association) **30** (2): 88–90. doi:10.2307/2683801. JSTOR 2683801.
- [2] Schwarz, W (2001). "The ex-Wald distribution as a descriptive model of response times". *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* **33** (4): 457–69. PMID 11816448.
- [3] Palmer, E. M.; Horowitz, T. S.; Torralba, A.; Wolfe, J. M. (2011). "What are the shapes of response time distributions in visual search?". *Journal of Experimental Psychology: Human Perception and Performance* **37**: 58. doi:10.1037/a0020747.
- [4] Schrodinger E (1915) Zur Theorie der Fall—und Steigversuche an Teilchenn mit Brownscher Bewegung. *Physikalische Zeitschrift* 16, 289-295
- [5] Folks, J. L.; Chhikara, R. S. (1978). "The Inverse Gaussian Distribution and Its Statistical Application--A Review". *Journal of the Royal Statistical Society. Series B (Methodological)* **40** (3): 263–289. doi:10.2307/2984691. JSTOR 2984691.
- [6] Giner, Goknur. "A monotonically convergent Newton iteration for the quantiles of any unimodal distribution, with application to the inverse Gaussian distribution" (PDF).

2.18.10 References

- *The inverse gaussian distribution: theory, methodology, and applications* by Raj Chhikara and Leroy Folks, 1989 ISBN 0-8247-7997-5
- *System Reliability Theory* by Marvin Rausand and Arnljot Høyland
- *The Inverse Gaussian Distribution* by Dr. V. Seshadri, Oxford Univ Press, 1993

2.18.11 External links

- [Inverse Gaussian Distribution](#) in Wolfram website.

2.19 Lévy distribution

For the more general family of Lévy alpha-stable distributions, of which this distribution is a special case, see [stable distribution](#).

In probability theory and statistics, the **Lévy distribution**, named after Paul Lévy, is a continuous probability distribution for a non-negative random variable. In spectroscopy, this distribution, with frequency as the dependent variable, is known as a **van der Waals profile**.^[note 1] It is a special case of the inverse-gamma distribution.

It is one of the few distributions that are stable and that have probability density functions that can be expressed analytically, the others being the normal distribution and the Cauchy distribution.

2.19.1 Definition

The probability density function of the Lévy distribution over the domain $x \geq \mu$ is

$$f(x; \mu, c) = \sqrt{\frac{c}{2\pi}} \frac{e^{-\frac{c}{2(x-\mu)}}}{(x-\mu)^{3/2}}$$

where μ is the location parameter and c is the scale parameter. The cumulative distribution function is

$$F(x; \mu, c) = \operatorname{erfc} \left(\sqrt{\frac{c}{2(x - \mu)}} \right)$$

where $\operatorname{erfc}(z)$ is the complementary **error function**. The shift parameter μ has the effect of shifting the curve to the right by an amount μ , and changing the support to the interval $[\mu, \infty)$. Like all **stable distributions**, the Lévy distribution has a standard form $f(x; 0, 1)$ which has the following property:

$$f(x; \mu, c) dx = f(y; 0, 1) dy$$

where y is defined as

$$y = \frac{x - \mu}{c}$$

The **characteristic function** of the Lévy distribution is given by

$$\varphi(t; \mu, c) = e^{i\mu t - \sqrt{-2ict}}.$$

Note that the characteristic function can also be written in the same form used for the stable distribution with $\alpha = 1/2$ and $\beta = 1$:

$$\varphi(t; \mu, c) = e^{i\mu t - |ct|^{1/2} (1 - i \operatorname{sign}(t))}.$$

Assuming $\mu = 0$, the n th **moment** of the unshifted Lévy distribution is formally defined by:

$$m_n \stackrel{\text{def}}{=} \sqrt{\frac{c}{2\pi}} \int_0^\infty \frac{e^{-c/2x} x^n}{x^{3/2}} dx$$

which diverges for all $n > 0$ so that the moments of the Lévy distribution do not exist. The **moment generating function** is then formally defined by:

$$M(t; c) \stackrel{\text{def}}{=} \sqrt{\frac{c}{2\pi}} \int_0^\infty \frac{e^{-c/2x+tx}}{x^{3/2}} dx$$

which diverges for $t > 0$ and is therefore not defined in an interval around zero, so that the moment generating function is not defined *per se*. Like all **stable distributions** except the **normal distribution**, the wing of the probability density function exhibits heavy tail behavior falling off according to a power law:

$$\lim_{x \rightarrow \infty} f(x; \mu, c) = \sqrt{\frac{c}{2\pi}} \frac{1}{x^{3/2}}.$$

This is illustrated in the diagram below, in which the probability density functions for various values of c and $\mu = 0$ are plotted on a log-log scale.

Differential equation

$$\left\{ f(x)(-c - 3\mu + 3x) + 2(x - \mu)^2 f'(x) = 0, f(0) = \frac{e^{\frac{c}{2\mu} \left(-\frac{c}{\mu} \right)^{3/2}}}{\sqrt{2\pi c}} \right\}$$

2.19.2 Related distributions

- If $X \sim \text{Levy}(\mu, c)$ then $kX + b \sim \text{Levy}(k\mu + b, kc)$
- If $X \sim \text{Levy}(0, c)$ then $X \sim \text{Inv-Gamma}(\frac{1}{2}, \frac{c}{2})$ (inverse gamma distribution)
- Lévy distribution is a special case of type 5 Pearson distribution
- If $Y \sim \text{Normal}(\mu, \sigma^2)$ (Normal distribution) then $(Y - \mu)^{-2} \sim \text{Levy}(0, 1/\sigma^2)$
- If $X \sim \text{Normal}(\mu, \frac{1}{\sqrt{\sigma}})$ then $(X - \mu)^{-2} \sim \text{Levy}(0, \sigma)$
- If $X \sim \text{Levy}(\mu, c)$ then $X \sim \text{Stable}(1/2, 1, c, \mu)$ (Stable distribution)
- If $X \sim \text{Levy}(0, c)$ then $X \sim \text{Scale-inv-}\chi^2(1, c)$ (Scaled-inverse-chi-squared distribution)
- If $X \sim \text{Levy}(\mu, c)$ then $(X - \mu)^{-\frac{1}{2}} \sim \text{FoldedNormal}(0, 1/\sqrt{c})$ (Folded normal distribution)

2.19.3 Random sample generation

Random samples from the Lévy distribution can be generated using **inverse transform sampling**. Given a random variate U drawn from the **uniform distribution** on the unit interval $(0, 1]$, the variate X given by

$$X = F^{-1}(U) = \frac{c}{(\Phi^{-1}(1 - U/2))^2} + \mu$$

is Lévy-distributed with location μ and scale c . Here $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

2.19.4 Applications

- The frequency of **geomagnetic reversals** appears to follow a Lévy distribution
- The **time of hitting** a single point α (different from the starting point 0) by the **Brownian motion** has the Lévy distribution with $c = \alpha^2$. (For a Brownian motion with drift, this time may follow an **inverse Gaussian distribution**, which has the Lévy distribution as a limit.)
- The length of the path followed by a photon in a turbid medium follows the Lévy distribution.^[1]
- A **Cauchy process** can be defined as a **Brownian motion subordinated** to a process associated with a Lévy distribution.^[2]

2.19.5 Footnotes

- [1] “van der Waals profile” appears with lowercase “van” in almost all sources, such as: *Statistical mechanics of the liquid surface* by Clive Anthony Croxton, 1980, A Wiley-Interscience publication, ISBN 0-471-27663-4, ISBN 978-0-471-27663-0, ; and in *Journal of technical physics*, Volume 36, by Instytut Podstawowych Problemów Techniki (Polska Akademia Nauk), publisher: Państwowe Wydawn. Naukowe., 1995,

2.19.6 Notes

- [1] Rogers, Geoffrey L, Multiple path analysis of reflectance from turbid media. *Journal of the Optical Society of America A*, **25**:11, p 2879-2883 (2008).
- [2] Applebaum, D. “Lectures on Lévy processes and Stochastic calculus, Braunschweig; Lecture 2: Lévy processes” (PDF). University of Sheffield. pp. 37–53.

2.19.7 References

- “[Information on stable distributions](#)”. Retrieved July 13, 2005. - John P. Nolan’s introduction to stable distributions, some papers on stable laws, and a free program to compute stable densities, cumulative distribution functions, quantiles, estimate parameters, etc. See especially [An introduction to stable distributions, Chapter 1](#)

2.19.8 External links

- Weisstein, Eric W., “[Lévy Distribution](#)”, *MathWorld*.
- [Lévy and stock prices](#)

2.20 Log-Cauchy distribution

In probability theory, a **log-Cauchy distribution** is a probability distribution of a random variable whose logarithm is distributed in accordance with a Cauchy distribution. If X is a random variable with a Cauchy distribution, then $Y = \exp(X)$ has a log-Cauchy distribution; likewise, if Y has a log-Cauchy distribution, then $X = \log(Y)$ has a Cauchy distribution.^[1]

2.20.1 Characterization

Probability density function

The log-Cauchy distribution has the probability density function:

$$\begin{aligned} f(x; \mu, \sigma) &= \frac{1}{x\pi\sigma \left[1 + \left(\frac{\ln x - \mu}{\sigma} \right)^2 \right]}, \quad x > 0 \\ &= \frac{1}{x\pi} \left[\frac{\sigma}{(\ln x - \mu)^2 + \sigma^2} \right], \quad x > 0 \end{aligned}$$

where μ is a real number and $\sigma > 0$.^{[1][2]} If σ is known, the scale parameter is e^μ .^[1] μ and σ correspond to the location parameter and scale parameter of the associated Cauchy distribution.^{[1][3]} Some authors define μ and σ as the location and scale parameters, respectively, of the log-Cauchy distribution.^[3]

For $\mu = 0$ and $\sigma = 1$, corresponding to a standard Cauchy distribution, the probability density function reduces to:^[4]

$$f(x; 0, 1) = \frac{1}{x\pi(1 + (\ln x)^2)}, \quad x > 0$$

Cumulative distribution function

The cumulative distribution function (cdf) when $\mu = 0$ and $\sigma = 1$ is:^[4]

$$F(x; 0, 1) = \frac{1}{2} + \frac{1}{\pi} \arctan(\ln x), \quad x > 0$$

Survival function

The **survival function** when $\mu = 0$ and $\sigma = 1$ is:^[4]

$$S(x; 0, 1) = \frac{1}{2} - \frac{1}{\pi} \arctan(\ln x), \quad x > 0$$

Hazard rate

The **hazard rate** when $\mu = 0$ and $\sigma = 1$ is:^[4]

$$\lambda(x; 0, 1) = \left(\frac{1}{x\pi(1 + (\ln x)^2)} \left(\frac{1}{2} - \frac{1}{\pi} \arctan(\ln x) \right) \right)^{-1}, \quad x > 0$$

The hazard rate decreases at the beginning and at the end of the distribution, but there may be an interval over which the hazard rate increases.^[4]

2.20.2 Properties

The log-Cauchy distribution is an example of a **heavy-tailed distribution**.^[5] Some authors regard it as a “super-heavy tailed” distribution, because it has a heavier tail than a **Pareto distribution**-type heavy tail, i.e., it has a **logarithmically decaying tail**.^{[5][6]} As with the Cauchy distribution, none of the non-trivial moments of the log-Cauchy distribution are finite.^[4] The mean is a moment so the log-Cauchy distribution does not have a defined mean or **standard deviation**.^{[7][8]}

The log-Cauchy distribution is **infinitely divisible** for some parameters but not for others.^[9] Like the **lognormal distribution**, **log-t or log-Student distribution** and **Weibull distribution**, the log-Cauchy distribution is a special case of the **generalized beta distribution of the second kind**.^{[10][11]} The log-Cauchy is actually a special case of the log-t distribution, similar to the Cauchy distribution being a special case of the **Student's t distribution** with 1 degree of freedom.^{[12][13]}

Since the Cauchy distribution is a **stable distribution**, the log-Cauchy distribution is a logstable distribution.^[14] Logstable distributions have **poles** at $x=0$.^[13]

2.20.3 Estimating parameters

The **median** of the natural logarithms of a sample is a **robust estimator** of μ .^[1] The **median absolute deviation** of the natural logarithms of a sample is a robust estimator of σ .^[1]

2.20.4 Uses

In **Bayesian statistics**, the log-Cauchy distribution can be used to approximate the **improper Jeffreys-Haldane density**, $1/k$, which is sometimes suggested as the **prior distribution** for k where k is a positive parameter being estimated.^{[15][16]} The log-Cauchy distribution can be used to model certain survival processes where significant **outliers** or extreme results may occur.^{[2][3][17]} An example of a process where a log-Cauchy distribution may be an appropriate model is the time between someone becoming infected with **HIV virus** and showing symptoms of the disease, which may be very long for some people.^[3] It has also been proposed as a model for species abundance patterns.^[18]

2.20.5 References

- [1] Olive, D.J. (June 23, 2008). “Applied Robust Statistics” (PDF). Southern Illinois University. p. 86. Retrieved 2011-10-18.

- [2] Lindsey, J.K. (2004). *Statistical analysis of stochastic processes in time*. Cambridge University Press. pp. 33, 50, 56, 62, 145. ISBN 978-0-521-83741-5.
- [3] Mode, C.J. & Sleeman, C.K. (2000). *Stochastic processes in epidemiology: HIV/AIDS, other infectious diseases*. World Scientific. pp. 29–37. ISBN 978-981-02-4097-4.
- [4] Marshall, A.W. & Olkin, I. (2007). *Life distributions: structure of nonparametric, semiparametric, and parametric families*. Springer. pp. 443–444. ISBN 978-0-387-20333-1.
- [5] Falk, M., Hüsler, J. & Reiss, R. (2010). *Laws of Small Numbers: Extremes and Rare Events*. Springer. p. 80. ISBN 978-3-0348-0008-2.
- [6] Alves, M.I.F., de Haan, L. & Neves, C. (March 10, 2006). “Statistical inference for heavy and super-heavy tailed distributions” (PDF).
- [7] “Moment”. Mathworld. Retrieved 2011-10-19.
- [8] Wang, Y. “Trade, Human Capital and Technology Spillovers: An Industry Level Analysis”. Carleton University. p. 14.
- [9] Bondesson, L. (2003). “On the Lévy Measure of the Lognormal and LogCauchy Distributions”. *Methodology and Computing in Applied Probability* (Kluwer Academic Publications): 243–256. Retrieved 2011-10-18.
- [10] Knight, J. & Satchell, S. (2001). *Return distributions in finance*. Butterworth-Heinemann. p. 153. ISBN 978-0-7506-4751-9.
- [11] Kemp, M. (2009). *Market consistency: model calibration in imperfect markets*. Wiley. ISBN 978-0-470-77088-7.
- [12] MacDonald, J.B. (1981). “Measuring Income Inequality”. In Taillie, C., Patil, G.P. & Baldessari, B. *Statistical distributions in scientific work: proceedings of the NATO Advanced Study Institute*. Springer. p. 169. ISBN 978-90-277-1334-6.
- [13] Kleiber, C. & Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Science*. Wiley. pp. 101–102, 110. ISBN 978-0-471-15064-0.
- [14] Panton, D.B. (May 1993). “Distribution function values for logstable distributions”. *Computers & Mathematics with Applications* **25** (9): 17–24. doi:10.1016/0898-1221(93)90128-I. Retrieved 2011-10-18.
- [15] Good, I.J. (1983). *Good thinking: the foundations of probability and its applications*. University of Minnesota Press. p. 102. ISBN 978-0-8166-1142-3.
- [16] Chen, M. (2010). *Frontiers of Statistical Decision Making and Bayesian Analysis*. Springer. p. 12. ISBN 978-1-4419-6943-9.
- [17] Lindsey, J.K., Jones, B. & Jarvis, P.; Jones; Jarvis (September 2001). “Some statistical issues in modelling pharmacokinetic data”. *Statistics in Medicine* **20** (17–18): 2775–278. doi:10.1002/sim.742. Retrieved 2011-10-19.
- [18] Zuo-Yun, Y.; et al. (June 2005). “LogCauchy, log-sech and lognormal distributions of species abundances in forest communities”. *Ecological Modelling* **184** (2–4): 329–340. doi:10.1016/j.ecolmodel.2004.10.011. Retrieved 2011-10-18.

2.21 Log-Laplace distribution

In probability theory and statistics, the **log-Laplace distribution** is the probability distribution of a random variable whose logarithm has a Laplace distribution. If X has a Laplace distribution with parameters μ and b , then $Y = e^X$ has a log-Laplace distribution. The distributional properties can be derived from the Laplace distribution.

2.21.1 Characterization

Probability density function

A random variable has a log-Laplace(μ, b) distribution if its probability density function is:^[1]

$$f(x|\mu, b) = \frac{1}{2bx} \exp\left(-\frac{|\ln x - \mu|}{b}\right)$$

$$= \frac{1}{2bx} \begin{cases} \exp\left(-\frac{\mu - \ln x}{b}\right) & \text{if } x < \mu \\ \exp\left(-\frac{\ln x - \mu}{b}\right) & \text{if } x \geq \mu \end{cases}$$

The cumulative distribution function for Y when $y > 0$, is

$$F(y) = 0.5 [1 + \operatorname{sgn}(\log(y) - \mu) (1 - \exp(-|\log(y) - \mu|/b))].$$

Versions of the log-Laplace distribution based on an asymmetric Laplace distribution also exist.^[2] Depending on the parameters, including asymmetry, the log-Laplace may or may not have a finite mean and a finite variance.^[2]

Differential equation

$$\begin{cases} \left\{ bx f'(x) + (b-1)f(x) = 0, f(1) = \frac{e^{-\frac{\mu}{b}}}{2b} \right\} & \text{if } x < \mu \\ \left\{ bx f'(x) + (b+1)f(x) = 0, f(1) = \frac{e^{\frac{\mu}{b}}}{2b} \right\} & \text{if } x \geq \mu \end{cases}$$

2.21.2 References

- [1] Lindsey, J.K. (2004). *Statistical analysis of stochastic processes in time*. Cambridge University Press. p. 33. ISBN 978-0-521-83741-5.
- [2] Kozubowski, T.J. & Podgorski, K. “A Log-Laplace Growth Rate Model” (PDF). University of Nevada-Reno. p. 4. Retrieved 2011-10-21.

2.22 Log-logistic distribution

In probability and statistics, the **log-logistic distribution** (known as the **Fisk distribution** in economics) is a continuous probability distribution for a non-negative random variable. It is used in survival analysis as a parametric model for events whose rate increases initially and decreases later, for example mortality rate from cancer following diagnosis or treatment. It has also been used in hydrology to model stream flow and precipitation, and in economics as a simple model of the distribution of wealth or income.

The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic distribution. It is similar in shape to the log-normal distribution but has heavier tails. Unlike the log-normal, its cumulative distribution function can be written in closed form.

2.22.1 Characterisation

There are several different parameterizations of the distribution in use. The one shown here gives reasonably interpretable parameters and a simple form for the cumulative distribution function.^{[1][2]} The parameter $\alpha > 0$ is a scale parameter and is also the median of the distribution. The parameter $\beta > 0$ is a shape parameter. The distribution is unimodal when $\beta > 1$ and its dispersion decreases as β increases.

The cumulative distribution function is

$$\begin{aligned}
F(x; \alpha, \beta) &= \frac{1}{1 + (x/\alpha)^{-\beta}} \\
&= \frac{(x/\alpha)^\beta}{1 + (x/\alpha)^\beta} \\
&= \frac{x^\beta}{\alpha^\beta + x^\beta}
\end{aligned}$$

where $x > 0$, $\alpha > 0$, $\beta > 0$.

The probability density function is

$$f(x; \alpha, \beta) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{(1 + (x/\alpha)^\beta)^2}$$

Alternative parameterization

An alternative parametrization is given by the pair μ, s in analogy with the logistic distribution:

$$\mu = \ln(\alpha)$$

$$s = 1/\beta$$

2.22.2 Properties

Moments

The k th raw moment exists only when $k < \beta$, when it is given by^{[3][4]}

$$\begin{aligned}
E(X^k) &= \alpha^k B(1 - k/\beta, 1 + k/\beta) \\
&= \alpha^k \frac{k \pi / \beta}{\sin(k \pi / \beta)}
\end{aligned}$$

where $B()$ is the beta function. Expressions for the mean, variance, skewness and kurtosis can be derived from this. Writing $b = \pi/\beta$ for convenience, the mean is

$$E(X) = \alpha b / \sin b, \quad \beta > 1,$$

and the variance is

$$\text{Var}(X) = \alpha^2 (2b / \sin 2b - b^2 / \sin^2 b), \quad \beta > 2.$$

Explicit expressions for the skewness and kurtosis are lengthy.^[5] As β tends to infinity the mean tends to α , the variance and skewness tend to zero and the excess kurtosis tends to 6/5 (see also related distributions below).

Quantiles

The **quantile function** (inverse cumulative distribution function) is :

$$F^{-1}(p; \alpha, \beta) = \alpha \left(\frac{p}{1-p} \right)^{1/\beta}.$$

It follows that the **median** is α , the lower **quartile** is $3^{1/\beta} \alpha$ and the upper quartile is $3^{-1/\beta} \alpha$.

2.22.3 Applications

Survival analysis

The log-logistic distribution provides one **parametric model** for **survival analysis**. Unlike the more commonly used **Weibull distribution**, it can have a non-monotonic **hazard function**: when $\beta > 1$, the hazard function is **unimodal** (when $\beta \leq 1$, the hazard decreases monotonically). The fact that the cumulative distribution function can be written in closed form is particularly useful for analysis of survival data with **censoring**.^[6] The log-logistic distribution can be used as the basis of an **accelerated failure time model** by allowing α to differ between groups, or more generally by introducing covariates that affect α but not β by modelling $\log(\alpha)$ as a linear function of the covariates.^[7]

The **survival function** is

$$S(t) = 1 - F(t) = [1 + (t/\alpha)^\beta]^{-1},$$

and so the **hazard function** is

$$h(t) = \frac{f(t)}{S(t)} = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{1 + (t/\alpha)^\beta}.$$

Hydrology

The log-logistic distribution has been used in hydrology for modelling stream flow rates and precipitation.^{[1][2]}

Extreme values like maximum one-day rainfall and river discharge per month or per year often follow a **log-normal distribution**.^[8] The log-normal distribution, however, needs a numeric approximation. As the log-logistic distribution, which can be solved analytically, is similar to the log-normal distribution, it can be used instead.

The blue picture illustrates an example of fitting the log-logistic distribution to ranked maximum one-day October rainfalls and it shows the 90% **confidence belt** based on the **binomial distribution**. The rainfall data are represented by the **plotting position** $r/(n+1)$ as part of the **cumulative frequency analysis**.

Economics

The log-logistic has been used as a simple model of the **distribution of wealth or income** in **economics**, where it is known as the Fisk distribution.^[9] Its **Gini coefficient** is $1/\beta$.^[10]

Networking

The log-logistic has been used as a model for the period of time beginning when some data leaves a software user application in a computer and the response is received by the same application after travelling through and being processed by

other computers, applications, and network segments, most or all of them without hard **real-time** guarantees (for example, when an application is displaying data coming from a remote **sensor** connected to the Internet). It has been shown to be a more accurate probabilistic model for that than the **log-normal distribution** or others, as long as abrupt changes of regime in the sequences of those times are properly detected.^[11]

2.22.4 Related distributions

- If $X \sim LL(\alpha, \beta)$ then $kX \sim LL(k\alpha, \beta)$
- $LL(\alpha, \beta) \sim \text{Dagum}(1, \alpha, \beta)$ (**Dagum distribution**)
- $LL(\alpha, \beta) \sim \text{SinghMaddala}(1, \alpha, \beta)$ (**Singh-Maddala distribution**)
- $LL(\gamma, \sigma) \sim \beta'(1, 1, \gamma, \sigma)$ **Beta prime distribution**
- If X has a log-logistic distribution with scale parameter α and shape parameter β then $Y = \log(X)$ has a **logistic distribution** with location parameter $\log(\alpha)$ and scale parameter $1/\beta$.
- As the shape parameter β of the log-logistic distribution increases, its shape increasingly resembles that of a (very narrow) **logistic distribution**. Informally, as $\beta \rightarrow \infty$,

$$LL(\alpha, \beta) \rightarrow L(\alpha, \alpha/\beta).$$

- The log-logistic distribution with shape parameter $\beta = 1$ and scale parameter α is the same as the **generalized Pareto distribution** with location parameter $\mu = 0$, shape parameter $\xi = 1$ and scale parameter α :

$$LL(\alpha, 1) = GPD(1, \alpha, 1).$$

- The addition of another parameter (a shift parameter) formally results in a **shifted log-logistic distribution**, but this is usually considered in a different parameterization so that the distribution can be bounded above or bounded below.

Generalizations

Several different distributions are sometimes referred to as the **generalized log-logistic distribution**, as they contain the log-logistic as a special case.^[10] These include the **Burr Type XII distribution** (also known as the *Singh-Maddala distribution*) and the **Dagum distribution**, both of which include a second shape parameter. Both are in turn special cases of the even more general *generalized beta distribution of the second kind*. Another more straightforward generalization of the log-logistic is the **shifted log-logistic distribution**.

2.22.5 See also

- **Probability distributions:** List of important distributions supported on semi-infinite intervals

2.22.6 References

- [1] Shoukri, M.M.; Mian, I.U.M.; Tracy, D.S. (1988), "Sampling Properties of Estimators of the Log-Logistic Distribution with Application to Canadian Precipitation Data", *The Canadian Journal of Statistics* **16** (3): 223–236, doi:10.2307/3314729, JSTOR 3314729
- [2] Ashkar, Fahim; Mahdi, Smail (2006), "Fitting the log-logistic distribution by generalized moments", *Journal of Hydrology* **328** (3–4): 694–703, doi:10.1016/j.jhydrol.2006.01.014

- [3] Tadikamalla, Pandu R.; Johnson, Norman L. (1982), “Systems of Frequency Curves Generated by Transformations of Logistic Variables”, *Biometrika* **69** (2): 461–465, doi:10.1093/biomet/69.2.461, JSTOR 2335422
- [4] Tadikamalla, Pandu R. (1980), “A Look at the Burr and Related Distributions”, *International Statistical Review* **48** (3): 337–344, doi:10.2307/1402945, JSTOR 1402945
- [5] McLaughlin, Michael P. (2001), *A Compendium of Common Probability Distributions* (PDF), p. A–37, retrieved 2008-02-15
- [6] Bennett, Steve (1983), “Log-Logistic Regression Models for Survival Data”, *Journal of the Royal Statistical Society, Series C* **32** (2): 165–171, doi:10.2307/2347295, JSTOR 2347295
- [7] Collett, Dave (2003), *Modelling Survival Data in Medical Research* (2nd ed.), CRC press, ISBN 1-58488-325-1
- [8] Ritzema (ed.), H.P. (1994), *Frequency and Regression Analysis* (PDF), Chapter 6 in: Drainage Principles and Applications, Publication 16, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands, pp. 175–224, ISBN 90-70754-33-9
- [9] Fisk, P.R. (1961), “The Graduation of Income Distributions”, *Econometrica* **29** (2): 171–185, doi:10.2307/1909287, JSTOR 1909287
- [10] Kleiber, C.; Kotz, S (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley, ISBN 0-471-15064-9
- [11] Gago-Benítez, A.; Fernández-Madrigal J.-A., Cruz-Martín, A. (2013), *Log-Logistic Modeling of Sensory Flow Delays in Networked Telerobots*, IEEE Sensors 13(8), pp. 2944 – 2953

2.23 Log-normal distribution

In probability theory, a **log-normal (or lognormal) distribution** is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then $Y = \ln(X)$ has a normal distribution. Likewise, if Y has a normal distribution, then $X = \exp(Y)$ has a log-normal distribution. A random variable which is log-normally distributed takes only positive real values. The distribution is occasionally referred to as the **Galton distribution** or **Galton’s distribution**, after Francis Galton.^[1] The log-normal distribution also has been associated with other names, such as McAlister, Gibrat and Cobb–Douglas.^[1]

A log-normal process is the statistical realization of the multiplicative product of many independent random variables, each of which is positive. This is justified by considering the central limit theorem in the log domain. The log-normal distribution is the maximum entropy probability distribution for a random variate X for which the mean and variance of $\ln(X)$ are specified.^[2]

2.23.1 Notation

Given a log-normally distributed random variable X and two parameters μ and σ that are, respectively, the mean and standard deviation of the variable’s natural logarithm, then the logarithm of X is normally distributed, and we can write X as

$$X = e^{\mu + \sigma Z}$$

with Z a standard normal variable.

This relationship is true regardless of the base of the logarithmic or exponential function. If $\log_a(Y)$ is normally distributed, then so is $\log_b(Y)$, for any two positive numbers $a, b \neq 1$. Likewise, if e^X is log-normally distributed, then so is a^X , where a is a positive number $\neq 1$.

On a logarithmic scale, μ and σ can be called the *location parameter* and the *scale parameter*, respectively.

In contrast, the mean, standard deviation, and variance of the non-logarithmized sample values are respectively denoted m , $s.d.$, and v in this article. The two sets of parameters can be related as (see also Arithmetic moments below)^[3]

$$\mu = \ln \left(\frac{m}{\sqrt{1 + \frac{v}{m^2}}} \right), \sigma = \sqrt{\ln \left(1 + \frac{v}{m^2} \right)}$$

2.23.2 Characterization

Probability density function

A random positive variable x is log-normally distributed if the logarithm of x is normally distributed,

$$\mathcal{N}(\ln x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right], \quad x > 0.$$

A change of variables must conserve differential probability. In particular,

$$\mathcal{N}(\ln x) d\ln x = \mathcal{N}(\ln x) \frac{d\ln x}{dx} dx = \mathcal{N}(\ln x) \frac{dx}{x} = \ln \mathcal{N}(x) dx,$$

where

$$\ln \mathcal{N}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right], \quad x > 0$$

is the log-normal probability density function.^[1]

Cumulative distribution function

The cumulative distribution function is

$$\int_0^x \ln \mathcal{N}(\xi; \mu, \sigma) d\xi = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\ln x - \mu}{\sigma\sqrt{2}} \right) \right] = \frac{1}{2} \operatorname{erfc} \left(-\frac{\ln x - \mu}{\sigma\sqrt{2}} \right) = \Phi \left(\frac{\ln x - \mu}{\sigma} \right),$$

where erfc is the complementary error function, and Φ is the cumulative distribution function of the standard normal distribution.

Characteristic function and moment generating function

All moments of the log-normal distribution exist and it holds that: $E[X^n] = e^{n\mu + \frac{n^2\sigma^2}{2}}$ (which can be derived by letting $z = \frac{\ln(x) - (\mu + n\sigma^2)}{\sigma}$ within the integral). However, the expected value $E[e^{tX}]$ is not defined for any positive value of the argument t as the defining integral diverges. In consequence the moment generating function is not defined.^[4] The last is related to the fact that the lognormal distribution is not uniquely determined by its moments.

Similarly, the characteristic function $E[e^{itX}]$ is not defined in the half complex plane and therefore it is not analytic in the origin. In consequence, the characteristic function of the log-normal distribution cannot be represented as an infinite convergent series.^[5] In particular, its Taylor formal series $\sum_{n=0}^{\infty} \frac{(it)^n}{n!} e^{n\mu + n^2\sigma^2/2}$ diverges. However, a number of alternative divergent series representations have been obtained^{[5][6][7][8]}

A closed-form formula for the characteristic function $\varphi(t)$ with t in the domain of convergence is not known. A relatively simple approximating formula is available in closed form and given by^[9]

$$\varphi(t) \approx \frac{\exp\left(-\frac{W^2(t\sigma^2 e^\mu) + 2W(t\sigma^2 e^\mu)}{2\sigma^2}\right)}{\sqrt{1+W(t\sigma^2 e^\mu)}}$$

where W is the **Lambert W function**. This approximation is derived via an asymptotic method but it stays sharp all over the domain of convergence of φ .

2.23.3 Properties

Location and scale

The location and scale parameters of a log-normal distribution, i.e. μ and σ , are more readily treated using the **geometric mean**, $\text{GM}[X]$, and the **geometric standard deviation**, $\text{GSD}[X]$, rather than the **arithmetic mean**, $\text{E}[X]$, and the **arithmetic standard deviation**, $\text{SD}[X]$.

Geometric moments The **geometric mean** of the log-normal distribution is $\text{GM}[X] = e^\mu$, and the **geometric standard deviation** is $\text{GSD}[X] = e^\sigma$.^{[10][11]} By analogy with the arithmetic statistics, one can define a **geometric variance**, $\text{GVar}[X] = e^{\sigma^2}$, and a **geometric coefficient of variation**,^[10] $\text{GCV}[X] = e^\sigma - 1$.

Because the **log-transformed** variable $Y = \ln X$ is symmetric and quantiles are preserved under monotonic transformations, the geometric mean of a log-normal distribution is equal to its median, $\text{Med}[X]$.^[12]

Note that the geometric mean is *less* than the arithmetic mean. This is due to the **AM–GM inequality**, and corresponds to the logarithm being convex down. In fact,

$$\text{E}[X] = e^{\mu + \frac{1}{2}\sigma^2} = e^\mu \cdot \sqrt{e^{\sigma^2}} = \text{GM}[X] \cdot \sqrt{\text{GVar}[X]}.$$

In finance the term $e^{-\frac{1}{2}\sigma^2}$ is sometimes interpreted as a **convexity correction**. From the point of view of **stochastic calculus**, this is the same correction term as in **Itô's lemma** for **geometric Brownian motion**.

Arithmetic moments The arithmetic mean, arithmetic variance, and arithmetic standard deviation of a log-normally distributed variable X are given by

$$\begin{aligned}\text{E}[X] &= e^{\mu + \frac{1}{2}\sigma^2}, \\ \text{Var}[X] &= (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} = (e^{\sigma^2} - 1)(\text{E}[X])^2, \\ \text{SD}[X] &= \sqrt{\text{Var}[X]} = e^{\mu + \frac{1}{2}\sigma^2} \sqrt{e^{\sigma^2} - 1} = \text{E}[X] \sqrt{e^{\sigma^2} - 1},\end{aligned}$$

respectively.

The location (μ) and scale (σ) parameters can be obtained if the arithmetic mean and the arithmetic variance are known; it is simpler if σ is computed first:

$$\begin{aligned}\mu &= \ln(\text{E}[X]) - \frac{1}{2} \ln\left(1 + \frac{\text{Var}[X]}{(\text{E}[X])^2}\right) = \ln(\text{E}[X]) - \frac{1}{2}\sigma^2, \\ \sigma^2 &= \ln\left(1 + \frac{\text{Var}[X]}{(\text{E}[X])^2}\right).\end{aligned}$$

For any real or complex number s , the s^{th} **moment** of a log-normally distributed variable X is given by^[11]

$$\text{E}[X^s] = e^{s\mu + \frac{1}{2}s^2\sigma^2}.$$

A log-normal distribution is not uniquely determined by its moments $E[X^k]$ for $k \geq 1$, that is, there exists some other distribution with the same moments for all k .^[1] In fact, there is a whole family of distributions with the same moments as the log-normal distribution.

Mode and median

The **mode** is the point of global maximum of the probability density function. In particular, it solves the equation $(\ln f)' = 0$:

$$\text{Mode}[X] = e^{\mu - \sigma^2}.$$

The **median** is such a point where $F_X = 0.5$:

$$\text{Med}[X] = e^{\mu}.$$

Arithmetic coefficient of variation

The arithmetic **coefficient of variation** $CV[X]$ is the ratio $\frac{SD[X]}{E[X]}$ (on the natural scale). For a log-normal distribution it is equal to

$$CV[X] = \sqrt{e^{\sigma^2} - 1}.$$

Contrary to the arithmetic standard deviation, the arithmetic coefficient of variation is independent of the arithmetic mean.

Partial expectation

The partial expectation of a random variable X with respect to a threshold k is defined as $g(k) = \int_k^\infty x \ln \mathcal{N}(x) dx$ where $\ln \mathcal{N}(x)$ is the probability density function of X . Alternatively, and using the definition of **conditional expectation**, it can be written as $g(k) = E[X|X > k]P(X > k)$. For a log-normal random variable the partial expectation is given by:

$$g(k) = \int_k^\infty x \ln \mathcal{N}(x) dx = e^{\mu + \frac{1}{2}\sigma^2} \Phi\left(\frac{\mu + \sigma^2 - \ln k}{\sigma}\right).$$

Where Phi is the **normal cumulative distribution function**. The derivation of the formula is provided in the discussion of this Wikipedia entry. The partial expectation formula has applications in insurance and economics, it is used in solving the partial differential equation leading to the **Black-Scholes formula**.

Conditional expectation

The conditional expectation of a lognormal random variable X with respect to a threshold k is its partial expectation divided by the cumulative probability of being in that range:

$$E[X|X < k] = e^{\mu + \sigma^2/2} \cdot \frac{\Phi\left[\frac{\ln(k) - \mu - \sigma^2}{\sigma}\right]}{\frac{1}{2} + \frac{1}{2} \operatorname{erf}\left[\frac{\ln(k) - \mu}{\sqrt{2}\sigma}\right]}$$

$$E[X|X \geq k] = e^{\mu + \sigma^2/2} \cdot \frac{\Phi\left[\frac{\mu + \sigma^2 - \ln(k)}{\sigma}\right]}{\frac{1}{2} - \frac{1}{2} \operatorname{erf}\left[\frac{\ln(k) - \mu}{\sqrt{2}\sigma}\right]}$$

Other

A set of data that arises from the log-normal distribution has a symmetric **Lorenz curve** (see also **Lorenz asymmetry coefficient**).^[13]

The harmonic H , geometric G and arithmetic A means of this distribution are related;^[14] such relation is given by

$$H = \frac{G^2}{A}.$$

Log-normal distributions are infinitely divisible,^[15] but they are not **stable distributions**, which can be easily drawn from.^[16]

2.23.4 Occurrence

The log-normal distribution is important in the description of natural phenomena. The reason is that for many natural processes of growth, relative **growth rate** is independent of size. This is also known as **Gibrat's law**, after Robert Gibrat (1904–1980) who formulated it for companies. It can be shown that a growth process following Gibrat's law will result in entity sizes with a log-normal distribution.^[17] Examples include:

- In **biology and medicine**,
 - Measures of size of living tissue (length, skin area, weight);^[18]
 - For highly communicable epidemics, such as SARS in 2003, if publication intervention is involved, the number of hospitalized cases is shown to satisfy the lognormal distribution with no free parameters if an entropy is assumed and the standard deviation is determined by the principle of maximum rate of entropy production.^[19]
 - The length of inert appendages (hair, claws, nails, teeth) of biological specimens, in the direction of growth;
 - Certain physiological measurements, such as blood pressure of adult humans (after separation on male/female subpopulations)^[20]

Consequently, **reference ranges** for measurements in healthy individuals are more accurately estimated by assuming a log-normal distribution than by assuming a symmetric distribution about the mean.

- In **colloidal chemistry and polymer chemistry**
 - Particle size distributions
 - Molar mass distributions
- In **hydrology**, the log-normal distribution is used to analyze extreme values of such variables as monthly and annual maximum values of daily rainfall and river discharge volumes.^[21]
 - The image on the right illustrates an example of fitting the log-normal distribution to ranked annually maximum one-day rainfalls showing also the 90% **confidence belt** based on the **binomial distribution**. The rainfall data are represented by **plotting positions** as part of a **cumulative frequency analysis**.
- In social sciences and demographics
 - In **economics**, there is evidence that the **income** of 97%–99% of the population is distributed log-normally.^[22] (The distribution of higher-income individuals follows a **Pareto distribution**.^[23])
 - In **finance**, in particular the Black–Scholes model, changes in the *logarithm* of exchange rates, price indices, and stock market indices are assumed normal^[24] (these variables behave like compound interest, not like simple interest, and so are multiplicative). However, some mathematicians such as **Benoît Mandelbrot** have argued^[25] that **log-Lévy distributions** which possesses **heavy tails** would be a more appropriate model, in particular for the analysis for stock market crashes. Indeed, stock price distributions typically exhibit a fat tail.^[26]

- city sizes
- Technology
 - In **reliability** analysis, the lognormal distribution is often used to model times to repair a maintainable system.^[27]
 - In **wireless communication**, “the local-mean power expressed in logarithmic values, such as dB or neper, has a normal (i.e., Gaussian) distribution.”^[28] Also, the random obstruction of radio signals due to large buildings and hills, called **shadowing**, is often modeled as a lognormal distribution.
 - It has been proposed that coefficients of friction and wear may be treated as having a lognormal distribution^[29]
 - In spray process, such as droplet impact, the size of secondary produced droplet has a lognormal distribution, with the standard deviation : $\sigma = \frac{\sqrt{6}}{6}$ determined by the principle of maximum rate of entropy production^[30] It is an open question whether this value of σ has some generality for other cases, though for spreading of communicable epidemics, σ is shown also to take this value.^[19]
 - Particle size distributions produced by comminution with random impacts, such as in **ball milling**
 - The **file size** distribution of publicly available audio and video data files (**MIME types**) follows a log-normal distribution over five **orders of magnitude**.^[31]
- The length of **chess** games tends to follow a log normal distribution.^[32]

2.23.5 Maximum likelihood estimation of parameters

For determining the **maximum likelihood** estimators of the log-normal distribution parameters μ and σ , we can use the same **procedure** as for the **normal distribution**. To avoid repetition, we observe that

$$L(x; \mu, \sigma) = \prod_{i=1}^n \left(\frac{1}{x_i} \right) \mathcal{N}(\ln x_i; \mu, \sigma)$$

where by L we denote the probability density function of the log-normal distribution and by $\mathcal{N}$ that of the normal distribution. Therefore, using the same indices to denote distributions, we can write the log-likelihood function thus:

$$\begin{aligned} \ell_L(\mu, \sigma | x_1, x_2, \dots, x_n) &= - \sum_k \ln x_k + \ell_N(\mu, \sigma | \ln x_1, \ln x_2, \dots, \ln x_n) \\ &= \text{constant} + \ell_N(\mu, \sigma | \ln x_1, \ln x_2, \dots, \ln x_n). \end{aligned}$$

Since the first term is constant with regard to μ and σ , both logarithmic likelihood functions, ℓ_L and ℓ_N , reach their maximum with the same μ and σ . Hence, using the formulas for the normal distribution maximum likelihood parameter estimators and the equality above, we deduce that for the log-normal distribution it holds that

$$\hat{\mu} = \frac{\sum_k \ln x_k}{n}, \hat{\sigma}^2 = \frac{\sum_k (\ln x_k - \hat{\mu})^2}{n}.$$

2.23.6 Multivariate log-normal

If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a **multivariate normal distribution** then $\mathbf{Y} = \exp(\mathbf{X})$ has a multivariate log-normal distribution^[33] with mean

$$\mathbb{E}[\mathbf{Y}]_i = e^{\mu_i + \frac{1}{2} \Sigma_{ii}},$$

and covariance matrix

$$\text{Var}[\mathbf{Y}]_{ij} = e^{\mu_i + \mu_j + \frac{1}{2}(\Sigma_{ii} + \Sigma_{jj})} (e^{\Sigma_{ij}} - 1).$$

2.23.7 Related distributions

- If $X \sim \mathcal{N}(\mu, \sigma^2)$ is a normal distribution, then $\exp(X) \sim \ln \mathcal{N}(\mu, \sigma^2)$.
- If $X \sim \ln \mathcal{N}(\mu, \sigma^2)$ is distributed log-normally, then $\ln(X) \sim \mathcal{N}(\mu, \sigma^2)$ is a normal random variable.
- If $X_j \sim \ln \mathcal{N}(\mu_j, \sigma_j^2)$ are n independent log-normally distributed variables, and $Y = \prod_{j=1}^n X_j$, then Y is also distributed log-normally:

$$Y \sim \ln \mathcal{N}\left(\sum_{j=1}^n \mu_j, \sum_{j=1}^n \sigma_j^2\right).$$

- Let $X_j \sim \ln \mathcal{N}(\mu_j, \sigma_j^2)$ be independent log-normally distributed variables with possibly varying σ and μ parameters, and $Y = \sum_{j=1}^n X_j$. The distribution of Y has no closed-form expression, but can be reasonably approximated by another log-normal distribution Z at the right tail.^[34] Its probability density function at the neighborhood of 0 has been characterized^[16] and it does not resemble any log-normal distribution. A commonly used approximation due to L.F. Fenton (but previously stated by R.I. Wilkinson and mathematical justified by Marlow^[35]) is obtained by matching the mean and variance of another lognormal distribution:

$$\sigma_Z^2 = \ln \left[\frac{\sum e^{2\mu_j + \sigma_j^2} (e^{\sigma_j^2} - 1)}{(\sum e^{\mu_j + \sigma_j^2/2})^2} + 1 \right],$$

$$\mu_Z = \ln \left[\sum e^{\mu_j + \sigma_j^2/2} \right] - \frac{\sigma_Z^2}{2}.$$

In the case that all X_j have the same variance parameter $\sigma_j = \sigma$, these formulas simplify to

$$\sigma_Z^2 = \ln \left[(e^{\sigma^2} - 1) \frac{\sum e^{2\mu_j}}{(\sum e^{\mu_j})^2} + 1 \right],$$

$$\mu_Z = \ln \left[\sum e^{\mu_j} \right] + \frac{\sigma^2}{2} - \frac{\sigma_Z^2}{2}.$$

- If $X \sim \ln \mathcal{N}(\mu, \sigma^2)$, then $X + c$ is said to have a *shifted log-normal* distribution with support $x \in (c, +\infty)$. $E[X + c] = E[X] + c$, $\text{Var}[X + c] = \text{Var}[X]$.
- If $X \sim \ln \mathcal{N}(\mu, \sigma^2)$, then $aX \sim \ln \mathcal{N}(\mu + \ln a, \sigma^2)$.
- If $X \sim \ln \mathcal{N}(\mu, \sigma^2)$, then $\frac{1}{X} \sim \ln \mathcal{N}(-\mu, \sigma^2)$.
- If $X \sim \ln \mathcal{N}(\mu, \sigma^2)$ then $X^a \sim \ln \mathcal{N}(a\mu, a^2\sigma^2)$. for $a \neq 0$
- Lognormal distribution is a special case of semi-bounded **Johnson distribution**
- If $X|Y \sim \text{Rayleigh}(Y)$ with $Y \sim \ln \mathcal{N}(\mu, \sigma^2)$, then $X \sim \text{Suzuki}(\mu, \sigma)$ (**Suzuki distribution**)

2.23.8 Similar distributions

A substitute for the log-normal whose integral can be expressed in terms of more elementary functions^[36] can be obtained based on the **logistic distribution** to get an approximation for the **CDF**

$$F(x; \mu, \sigma) = \left[\left(\frac{e^\mu}{x} \right)^{\pi/(\sigma\sqrt{3})} + 1 \right]^{-1}.$$

This is a **log-logistic distribution**.

2.23.9 See also

- **Log-distance path loss model**
- **Slow fading**

2.23.10 Notes

- [1] Johnson, Norman L.; Kotz, Samuel; Balakrishnan, N. (1994), "14: Lognormal Distributions", *Continuous univariate distributions. Vol. 1*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (2nd ed.), New York: John Wiley & Sons, ISBN 978-0-471-58495-7, MR 1299979
- [2] Park, Sung Y.; Bera, Anil K. (2009). "Maximum entropy autoregressive conditional heteroskedasticity model" (PDF). *Journal of Econometrics* (Elsevier) **150** (2): 219–230. doi:10.1016/j.jeconom.2008.12.014. Retrieved 2011-06-02.
- [3] "Lognormal mean and variance"
- [4] Heyde, CC. (1963), "On a property of the lognormal distribution", *Journal of the Royal Statistical Society, Series B (Methodological)* **25** (2): 392–393, doi:10.1007/978-1-4419-5823-5_6
- [5] Holgate, P. (1989). "The lognormal characteristic function, vol. 18, pp. 4539–4548, 1989". *Communications in Statistical – Theory and Methods* **18** (12): 4539–4548. doi:10.1080/03610928908830173.
- [6] Barakat, R. (1976). "Sums of independent lognormally distributed random variables". *Journal of the Optical Society of America* **66** (3): 211–216. doi:10.1364/JOSA.66.000211.
- [7] Barouch, E.; Kaufman, GM.; Glasser, ML. (1986). "On sums of lognormal random variables" (PDF). *Studies in Applied Mathematics* **75** (1): 37–55.
- [8] Leipnik, Roy B. (January 1991). "On Lognormal Random Variables: I – The Characteristic Function". *Journal of the Australian Mathematical Society Series B* **32** (3): 327–347. doi:10.1017/S0334270000006901.
- [9] S. Asmussen, J.L. Jensen, L. Rojas-Nandayapa. "On the Laplace transform of the Lognormal distribution", Thiele centre preprint, (2013).
- [10] Kirkwood, Thomas BL (Dec 1979). "Geometric means and measures of dispersion". *Biometrics* **35** (4): 908–9. doi:10.2307/2530139.
- [11] Limpert, E; Stahel, W; Abbt, M (2001). "Lognormal distributions across the sciences: keys and clues". *BioScience* **51** (5): 341–352. doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.
- [12] Daly, Leslie E.; Bourke, Geoffrey Joseph (2000). *Interpretation and uses of medical statistics* (5th ed.). Wiley-Blackwell. p. 89. doi:10.1002/9780470696750. ISBN 978-0-632-04763-5.
- [13] Damgaard, Christian; Weiner, Jacob (2000). "Describing inequality in plant size or fecundity". *Ecology* **81** (4): 1139–1142. doi:10.1890/0012-9658(2000)081[1139:DIIPSO]2.0.CO;2.
- [14] Rossman, Lewis A (July 1990). "Design stream flows based on harmonic means". *J Hydraulic Engineering* **116** (7): 946–950. doi:10.1061/(ASCE)0733-9429(1990)116:7(946).

- [15] Olof Thorin (1977), ["On the Infinite Divisibility of the Log Normal Distribution"]. *Scandinavian Actuarial Journal*, 03/1977; 1977(3):121-148. Template:10.1080/03461238.1977.10405635
- [16] Gao, X.; Xu, H; Ye, D. (2009), "Asymptotic Behaviors of Tail Density for Sum of Correlated Lognormal Variables". *International Journal of Mathematics and Mathematical Sciences*, vol. 2009, Article ID 630857. doi:10.1155/2009/630857
- [17] Sutton, John (Mar 1997). "Gibrat's Legacy". *Journal of Economic Literature* **32** (1): 40–59. JSTOR 2729692.
- [18] Huxley, Julian S. (1932). *Problems of relative growth*. London. ISBN 0-486-61114-0. OCLC 476909537.
- [19] Wang, WB; Wang, CF; Wu, ZN; Hu, RF (2013). "Modelling the spreading rate of controlled communicable epidemics through an entropy-based thermodynamic model" **56** (11). *SCIENCE CHINA Physics, Mechanics & Astronomy*. pp. 2143–2150.
- [20] Makuch, Robert W.; D.H. Freeman; M.F. Johnson (1979). "Justification for the lognormal distribution as a model for blood pressure". *Journal of Chronic Diseases* **32** (3): 245–250. doi:10.1016/0021-9681(79)90070-5. Retrieved 27 February 2012.
- [21] Ritzema (ed.), H.P. (1994). *Frequency and Regression Analysis* (PDF). Chapter 6 in: *Drainage Principles and Applications*, Publication 16, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. pp. 175–224. ISBN 90-70754-33-9.
- [22] Clementi, Fabio; Gallegati, Mauro (2005) "Pareto's law of income distribution: Evidence for Germany, the United Kingdom, and the United States", EconWPA
- [23] Wataru, Souma (2002-02-22). "Physics of Personal Income". arXiv:cond-mat/0202388.
- [24] Black, F.; Scholes, M. (1973). "The Pricing of Options and Corporate Liabilities". *Journal of Political Economy* **81** (3): 637. doi:10.1086/260062.
- [25] Mandelbrot, Benoit (2004). *The (mis-)Behaviour of Markets*. Basic Books. ISBN 9780465043552.
- [26] Bunchen, P., *Advanced Option Pricing*, University of Sydney coursebook, 2007
- [27] O'Connor, Patrick; Kleyner, Andre (2011). *Practical Reliability Engineering*. John Wiley & Sons. p. 35. ISBN 978-0-470-97982-2.
- [28] <http://wireless.per.nl/reference/chaptr03/shadow/shadow.htm> Archived May 9, 2015 at the Wayback Machine
- [29] Steele, C. (2008). "Use of the lognormal distribution for the coefficients of friction and wear". *Reliability Engineering & System Safety* **93** (10): 1574–2013. doi:10.1016/j.ress.2007.09.005.
- [30] Wu, Zi-Niu (July 2003). "Prediction of the size distribution of secondary ejected droplets by crown splashing of droplets impinging on a solid wall". *Probabilistic Engineering Mechanics* **18** (3): 241–249. doi:10.1016/S0266-8920(03)00028-6.
- [31] Gros, C; Kaczor, G.; Markovic, D (2012). "Neuropsychological constraints to human data production on a global scale". *The European Physical Journal B* **85** (28). doi:10.1140/epjb/e2011-20581-3.
- [32] <http://chess.stackexchange.com/questions/2506/what-is-the-average-length-of-a-game-of-chess/4899#4899>
- [33] Tarmast, Ghasem (2001). *Multivariate Log-Normal Distribution* (PDF). ISI Proceedings: 53rd Session. Seoul.
- [34] Asmussen, S.; Rojas-Nandayapa, L. (2008). "Asymptotics of Sums of Lognormal Random Variables with Gaussian Copula". *Statistics and Probability Letters* **78** (16): 2709–2714. doi:10.1016/j.spl.2008.03.035.
- [35] Marlow, NA. (Nov 1967). "A normal limit theorem for power sums of independent normal random variables". *Bell System Technical Journal* **46** (9): 2081–2089. doi:10.1002/j.1538-7305.1967.tb04244.x.
- [36] Swamee, P. K. (2002). "Near Lognormal Distribution". *Journal of Hydrologic Engineering* **7** (6): 441–444. doi:10.1061/(ASCE)1084-0699(2002)7:6(441).

2.23.11 References

- Crow, Edwin L.; Shimizu, Kunio (Editors) (1988), *Lognormal Distributions, Theory and Applications*, Statistics: Textbooks and Monographs **88**, New York: Marcel Dekker, Inc., pp. xvi+387, ISBN 0-8247-7803-0, MR 0939191, Zbl 0644.62014
- Aitchison, J. and Brown, J.A.C. (1957) *The Lognormal Distribution*, Cambridge University Press.
- E. Limpert, W. Stahel and M. Abbt (2001) *Log-normal Distributions across the Sciences: Keys and Clues*, BioScience, 51 (5), 341–352.
- Eric W. Weisstein et al. Log Normal Distribution at MathWorld. Electronic document, retrieved October 26, 2006.
- Holgate, P. (1989). “The lognormal characteristic function”. *Communications in Statistics - Theory and Methods* **18** (12): 4539–4548. doi:10.1080/03610928908830173.

2.23.12 Further reading

- Brooks, Robert; Corson, Jon; Donal, Wales (1994). “The Pricing of Index Options When the Underlying Assets All Follow a Lognormal Diffusion”. *Advances in Futures and Options Research* **7**.

2.23.13 External links

2.24 Lomax distribution

The **Lomax distribution**, conditionally also called the **Pareto Type II distribution**, is a heavy-tail probability distribution often used in business, economics, and actuarial modeling.^{[1][2]} It is named after K. S. Lomax. It is essentially a **Pareto distribution** that has been shifted so that its support begins at zero.^[3]

2.24.1 Characterization

Probability density function

The **probability density function** (pdf) for the Lomax distribution is given by

$$p(x) = \frac{\alpha}{\lambda} \left[1 + \frac{x}{\lambda} \right]^{-(\alpha+1)}, \quad x \geq 0,$$

with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$. The density can be rewritten in such a way that more clearly shows the relation to the **Pareto Type I distribution**. That is:

$$p(x) = \frac{\alpha \lambda^\alpha}{(x + \lambda)^{\alpha+1}}$$

Differential equation

The pdf of the Lomax distribution is a solution to the following **differential equation**:

$$\left\{ \begin{array}{l} (\lambda + x)p'(x) + (\alpha + 1)p(x) = 0, \\ p(0) = \frac{\alpha}{\lambda} \end{array} \right\}$$

2.24.2 Relation to the Pareto distribution

The Lomax distribution is a **Pareto Type I distribution** shifted so that its support begins at zero. Specifically:

If $Y \sim \text{Pareto}(x_m = \lambda, \alpha)$, then $Y - x_m \sim \text{Lomax}(\lambda, \alpha)$.

The Lomax distribution is a **Pareto Type II distribution** with $x_m = \lambda$ and $\mu = 0$:^[4]

If $X \sim \text{Lomax}(\lambda, \alpha)$ then $X \sim \text{P(II)}(x_m = \lambda, \alpha, \mu = 0)$.

2.24.3 Relation to generalized Pareto distribution

The Lomax distribution is a special case of the **generalized Pareto distribution**. Specifically:

$$\mu = 0, \xi = \frac{1}{\alpha}, \sigma = \frac{\lambda}{\alpha}.$$

2.24.4 Relation to q-exponential distribution

The Lomax distribution is a special case of the **q-exponential distribution**. The q-exponential extends this distribution to support on a bounded interval. The Lomax parameters are given by:

$$\alpha = \frac{2-q}{q-1}, \lambda = \frac{1}{\lambda_q(q-1)}.$$

2.24.5 Non-central moments

The ν th non-central moment $E[X^\nu]$ exists only if the shape parameter α strictly exceeds ν , when the moment has the value

$$E(X^\nu) = \frac{\lambda^\nu \Gamma(\alpha - \nu) \Gamma(1 + \nu)}{\Gamma(\alpha)}$$

2.24.6 See also

- **Power law**

2.24.7 References

- [1] Lomax, K. S. (1954) "Business Failures; Another example of the analysis of failure data". *Journal of the American Statistical Association*, 49, 847–852. JSTOR 2281544
- [2] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994) *Continuous Univariate Distributions, Volume 1*, 2nd Edition, Wiley. ISBN 0-471-58495-9 (pages 575, 602)
- [3] Van Hauwermeiren M and Vose D (2009). *A Compendium of Distributions* [ebook]. Vose Software, Ghent, Belgium. Available at www.vosesoftware.com. Accessed 07/07/11
- [4] Kleiber, Christian; Kotz, Samuel (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley Series in Probability and Statistics **470**, John Wiley & Sons, p. 60, ISBN 9780471457169.

2.25 Geometric stable distribution

A **geometric stable distribution** or **geo-stable distribution** is a type of **leptokurtic probability distribution**. Geometric stable distributions were introduced in Klebanov, L. B., Maniya, G. M., and Melamed, I. A. (1985). A problem of Zolotarev and analogs of infinitely divisible and stable distributions in a scheme for summing a random number of random variables, *Theory of Probability & Its Applications*, 29(4):791–794. These distributions are analogues for stable distributions for the case when the number of summands is random, independent on the distribution of summand and having geometric distribution. The geometric stable distribution may be symmetric or asymmetric. A symmetric geometric stable distribution is also referred to as a **Linnik distribution**. The **Laplace distribution** is a special case of the geometric stable distribution and of a Linnik distribution. The **Mittag-Leffler distribution** is also a special case of a geometric stable distribution.

The geometric stable distribution has applications in finance theory.^{[1][2][3]}

2.25.1 Characteristics

For most geometric stable distributions, the **probability density function** and **cumulative distribution function** have no closed form solution. But a geometric stable distribution can be defined by its **characteristic function**, which has the form:^[4]

$$\varphi(t; \alpha, \beta, \lambda, \mu) = [1 + \lambda^\alpha |t|^\alpha \omega - i\mu t]^{-1}$$

$$\text{where } \omega = \begin{cases} 1 - i \tan \frac{\pi\alpha}{2} \beta \operatorname{sign}(t) & \text{if } \alpha \neq 1 \\ 1 + i \frac{2}{\pi} \beta \log |t| \operatorname{sign}(t) & \text{if } \alpha = 1 \end{cases}$$

α , which must be greater than 0 and less than or equal to 2, is the shape parameter or index of stability, which determines how heavy the tails are.^[4] Lower α corresponds to **heavier tails**.

β , which must be greater than or equal to -1 and less than or equal to 1 , is the skewness parameter.^[4] When β is negative the distribution is skewed to the left and when β is positive the distribution is skewed to the right. When β is zero the distribution is symmetric, and the characteristic function reduces to:^[4]

$$\varphi(t; \alpha, 0, \lambda, \mu) = [1 + \lambda^\alpha |t|^\alpha - i\mu t]^{-1}$$

The symmetric geometric stable distribution with $\mu = 0$ is also referred to as a Linnik distribution.^{[5][6]} A completely skewed geometric stable distribution, that is with $\beta = 1$, $\alpha < 1$, with $0 < \mu < 1$ is also referred to as a Mittag-Leffler distribution.^[7] Although β determines the skewness of the distribution, it should not be confused with the typical **skewness coefficient** or **3rd standardized moment**, which in most circumstances is undefined for a geometric stable distribution.

$\lambda > 0$ is the **scale parameter** and μ is the location parameter.^[4]

When $\alpha = 2$, $\beta = 0$ and $\mu = 0$ (i.e., a symmetric geometric stable distribution or Linnik distribution with $\alpha = 2$), the distribution becomes the symmetric **Laplace distribution** with mean of 0,^[5] which has a **probability density function** of:

$$f(x|0, \lambda) = \frac{1}{2\lambda} \exp\left(-\frac{|x|}{\lambda}\right)$$

The Laplace distribution has a **variance** equal to $2\lambda^2$. However, for $\alpha < 2$ the variance of the geometric stable distribution is infinite.

2.25.2 Relationship to the stable distribution

The **stable distribution** has the property that if $X_1, X_2, \dots, X_n$ are independent, identically distributed random variables taken from a stable distribution, the sum $Y = a_n(X_1 + X_2 + \dots + X_n) + b_n$ has the same distribution as the X_i s for

some a_n and b_n .

The geometric stable distribution has a similar property, but where the number of elements in the sum is a **geometrically distributed** random variable. If $X_1, X_2, \dots$ are **independent and identically distributed random variables** taken from a geometric stable distribution, the **limit** of the sum $Y = a_{N_p}(X_1 + X_2 + \dots + X_{N_p}) + b_{N_p}$ approaches the distribution of the X_i s for some coefficients a_{N_p} and b_{N_p} as p approaches 0, where N_p is a random variable independent of the X_i s taken from a geometric distribution with parameter p .^[2] In other words:

$$\Pr(N_p = n) = (1 - p)^{n-1} p.$$

The distribution is strictly geometric stable only if the sum $Y = a(X_1 + X_2 + \dots + X_{N_p})$ equals the distribution of the X_i s for some a .^[1]

There is also a relationship between the stable distribution characteristic function and the geometric stable distribution characteristic function. The stable distribution has a characteristic function of the form:

$$\Phi(t; \alpha, \beta, \lambda, \mu) = \exp [it\mu - |\lambda t|^\alpha (1 - i\beta \operatorname{sign}(t)\Omega)],$$

where

$$\Omega = \begin{cases} \tan \frac{\pi\alpha}{2} & \text{if } \alpha \neq 1, \\ -\frac{2}{\pi} \log |t| & \text{if } \alpha = 1. \end{cases}$$

The geometric stable characteristic function can be expressed in terms of a stable characteristic function as:^[8]

$$\varphi(t; \alpha, \beta, \lambda, \mu) = [1 - \log(\Phi(t; \alpha, \beta, \lambda, \mu))]^{-1}.$$

2.25.3 References

- [1] Rachev, S. & Mittnik, S. (2000). *Stable Paretian Models in Finance*. Wiley. pp. 34–36. ISBN 978-0-471-95314-2.
- [2] Trindade, A.A.; Zhu, Y. & Andrews, B. (May 18, 2009). “Time Series Models With Asymmetric Laplace Innovations” (PDF). pp. 1–3. Retrieved 2011-02-27.
- [3] Meerschaert, M. & Scejfler, H. “Limit Theorems for Continuous Time Random Walks” (PDF). p. 15. Retrieved 2011-02-27.
- [4] Kozubowski, T.; Podgorski, K. & Samorodnitsky, G. “Tails of Lévy Measure of Geometric Stable Random Variables” (PDF). pp. 1–3. Retrieved 2011-02-27.
- [5] Kotz, S.; Kozubowski, T. & Podgórski, K. (2001). *The Laplace distribution and generalizations*. Birkhäuser. pp. 199–200. ISBN 978-0-8176-4166-5.
- [6] Kozubowski, T. (2006). “A Note on Certain Stability and Limiting Properties of ν -infinitely divisible distribution” (PDF). *Int. J. Contemp. Math. Sci.* **1** (4): 159. Retrieved 2011-02-27.
- [7] Burnecki, K.; Janczura, J.; Magdziarz, M. & Weron, A. (2008). “Can One See a Competition Between Subdiffusion and Lévy Flights? A Care of Geometric Stable Noise” (PDF). *Acta Physica Polonica B* **39** (8): 1048. Retrieved 2011-02-27.
- [8] “Geometric Stable Laws Through Series Representations” (PDF). *Serdica Mathematical Journal* **25**: 243. 1999. Retrieved 2011-02-28.

2.26 Nakagami distribution

The **Nakagami distribution** or the **Nakagami- m distribution** is a probability distribution related to the gamma distribution. It has two parameters: a shape parameter m and a second parameter controlling spread, Ω .

2.26.1 Characterization

Its probability density function (pdf) is^[1]

$$f(x; m, \Omega) = \frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} \exp\left(-\frac{m}{\Omega} x^2\right).$$

Its cumulative distribution function is^[1]

$$F(x; m, \Omega) = P\left(m, \frac{m}{\Omega} x^2\right)$$

where P is the incomplete gamma function (regularized).

Differential equation

$$\left\{ x\Omega f'(x) + f(x) (2mx^2 - 2m\Omega + \Omega) = 0, f(1) = \frac{2m^m e^{-\frac{m}{\Omega}} \Omega^{-m}}{\Gamma(m)} \right\}$$

2.26.2 Parameter estimation

The parameters m and Ω are^[2]

$$m = \frac{E^2[X^2]}{\text{Var}[X^2]},$$

and

$$\Omega = E[X^2].$$

An alternative way of fitting the distribution is to re-parametrize Ω and m as $\sigma = \Omega/m$ and m .^[3] Then, by taking the derivative of log likelihood with respect to each of the new parameters, the following equations are obtained and these can be solved using the **Newton-Raphson** method:

$$\Gamma(m) = \frac{x^{2m}}{\sigma^m},$$

and

$$\sigma = \frac{x^2}{m}$$

It is reported by authors that modelling data with Nakagami distribution and estimating parameters by above mention method results in better performance for low data regime compared to moments based methods.

2.26.3 Generation

The Nakagami distribution is related to the **gamma distribution**. In particular, given a random variable $Y \sim \text{Gamma}(k, \theta)$, it is possible to obtain a random variable $X \sim \text{Nakagami}(m, \Omega)$, by setting $k = m$, $\theta = \Omega/m$, and taking the square root of Y :

$$X = \sqrt{Y}$$

The Nakagami distribution $f(y; m, \Omega)$ can be generated from the **chi distribution** with parameter k set to $2m$ and then following it by a scaling transformation of random variables. That is, a Nakagami random variable X is generated by a simple scaling transformation on a Chi-distributed random variable $Y \sim \chi(2m)$ as below:

$$X = \sqrt{(\Omega/2m)} Y.$$

2.26.4 History and applications

The Nakagami distribution is relatively new, being first proposed in 1960.^[4] It has been used to model attenuation of wireless signals traversing multiple paths.^[5]

2.26.5 References

- [1] Laurenson, Dave (1994). “Nakagami Distribution”. *Indoor Radio Channel Propagation Modelling by Ray Tracing Techniques*. Retrieved 2007-08-04.
- [2] R. Kolar, R. Jirik, J. Jan (2004) “Estimator Comparison of the Nakagami-m Parameter and Its Application in Echocardiography”, *Radioengineering*, 13 (1), 8–12
- [3] Mitra, Rangeet; Mishra, Amit Kumar; Choubisa, Tarun (2012). “Maximum Likelihood Estimate of Parameters of Nakagami-m Distribution”. *International Conference on Communications, Devices and Intelligent Systems (CODIS)*, 2012: 9-12.
- [4] Nakagami, M. (1960) “The m-Distribution, a general formula of intensity of rapid fading”. In William C. Hoffman, editor, *Statistical Methods in Radio Wave Propagation: Proceedings of a Symposium held June 18-20, 1958*, pp 3-36. Pergamon Press.
- [5] Parsons, J. D. (1992) *The Mobile Radio Propagation Channel*. New York: Wiley.

2.27 Pareto distribution

The **Pareto distribution**, named after the Italian civil engineer, economist, and sociologist Vilfredo Pareto, is a power law probability distribution that is used in description of social, scientific, geophysical, actuarial, and many other types of observable phenomena.

2.27.1 Definition

If X is a **random variable** with a Pareto (Type I) distribution,^[1] then the probability that X is greater than some number x , i.e. the **survival function** (also called tail function), is given by

$$\bar{F}(x) = \Pr(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^\alpha & x \geq x_m, \\ 1 & x < x_m. \end{cases}$$

where x_m is the (necessarily positive) minimum possible value of X , and α is a positive parameter. The Pareto Type I distribution is characterized by a scale parameter x_m and a shape parameter α , which is known as the *tail index*. When this distribution is used to model the distribution of wealth, then the parameter α is called the **Pareto index**.

2.27.2 Properties

Cumulative distribution function

From the definition, the cumulative distribution function of a Pareto random variable with parameters α and x_m is

$$F_X(x) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^\alpha & x \geq x_m, \\ 0 & x < x_m. \end{cases}$$

When plotted on linear axes, the distribution assumes the familiar J-shaped curve which approaches each of the orthogonal axes *asymptotically*. All segments of the curve are self-similar (subject to appropriate scaling factors). When plotted in a *log-log plot*, the distribution is represented by a straight line.

Probability density function

It follows (by differentiation) that the probability density function is

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & x \geq x_m, \\ 0 & x < x_m. \end{cases}$$

Moments and characteristic function

- The expected value of a random variable following a Pareto distribution is

$$E(X) = \begin{cases} \infty & \alpha \leq 1, \\ \frac{\alpha x_m}{\alpha - 1} & \alpha > 1. \end{cases}$$

- The variance of a random variable following a Pareto distribution is

$$\text{Var}(X) = \begin{cases} \infty & \alpha \in (1, 2], \\ \left(\frac{x_m}{\alpha - 1}\right)^2 \frac{\alpha}{\alpha - 2} & \alpha > 2. \end{cases}$$

(If $\alpha \leq 1$, the variance does not exist.)

- The raw moments are

$$\mu'_n = \begin{cases} \infty & \alpha \leq n, \\ \frac{\alpha x_m^n}{\alpha - n} & \alpha > n. \end{cases}$$

- The moment generating function is only defined for non-positive values $t \leq 0$ as

$$M(t; \alpha, x_m) = E[e^{tX}] = \alpha(-x_m t)^\alpha \Gamma(-\alpha, -x_m t)$$

$$M(0, \alpha, x_m) = 1.$$

- The characteristic function is given by

$$\varphi(t; \alpha, x_m) = \alpha(-ix_m t)^\alpha \Gamma(-\alpha, -ix_m t),$$

where $\Gamma(a, x)$ is the incomplete gamma function.

Conditional distributions

The conditional probability distribution of a Pareto-distributed random variable, given the event that it is greater than or equal to a particular number x_1 exceeding x_m , is a Pareto distribution with the same Pareto index α but with minimum x_1 instead of x_m .

A characterization theorem

Suppose $X_1, X_2, X_3, \dots$ are independent identically distributed random variables whose probability distribution is supported on the interval $[x_m, \infty)$ for some $x_m > 0$. Suppose that for all n , the two random variables $\min\{X_1, \dots, X_n\}$ and $(X_1 + \dots + X_n) / \min\{X_1, \dots, X_n\}$ are independent. Then the common distribution is a Pareto distribution.

Geometric mean

The geometric mean (G) is^[2]

$$G = x_m \exp\left(\frac{1}{\alpha}\right)$$

Harmonic mean

The harmonic mean (H) is^[2]

$$H = x_m \left(1 + \frac{1}{\alpha}\right)$$

2.27.3 Generalized Pareto distributions

See also: Generalized Pareto distribution

There is a hierarchy^{[1][3]} of Pareto distributions known as Pareto Type I, II, III, IV, and Feller–Pareto distributions.^{[1][3][4]} Pareto Type IV contains Pareto Type I–III as special cases. The Feller–Pareto^{[3][5]} distribution generalizes Pareto Type IV.

Pareto types I–IV

The Pareto distribution hierarchy is summarized in the next table comparing the **survival functions** (complementary CDF).

When $\mu = 0$, the Pareto distribution Type II is also known as the **Lomax distribution**.^[6]

In this section, the symbol x_m , used before to indicate the minimum value of x , is replaced by σ .

The shape parameter α is the tail index, μ is location, σ is scale, γ is an inequality parameter. Some special cases of Pareto Type (IV) are

$$P(IV)(\sigma, \sigma, 1, \alpha) = P(I)(\sigma, \alpha),$$

$$P(IV)(\mu, \sigma, 1, \alpha) = P(II)(\mu, \sigma, \alpha),$$

$$P(IV)(\mu, \sigma, \gamma, 1) = P(III)(\mu, \sigma, \gamma).$$

The finiteness of the mean, and the existence and the finiteness of the variance depend on the tail index α (inequality index γ). In particular, fractional δ -moments are finite for some $\delta > 0$, as shown in the table below, where δ is not necessarily an integer.

Feller–Pareto distribution

Feller^{[3][5]} defines a Pareto variable by transformation $U = Y^{-1} - 1$ of a **beta random variable** Y , whose probability density function is

$$f(y) = \frac{y^{\gamma_1-1}(1-y)^{\gamma_2-1}}{B(\gamma_1, \gamma_2)}, \quad 0 < y < 1; \gamma_1, \gamma_2 > 0,$$

where $B(\cdot)$ is the **beta function**. If

$$W = \mu + \sigma(Y^{-1} - 1)^\gamma, \quad \sigma > 0, \gamma > 0,$$

then W has a Feller–Pareto distribution $FP(\mu, \sigma, \gamma, \gamma_1, \gamma_2)$.^[1]

If $U_1 \sim \Gamma(\delta_1, 1)$ and $U_2 \sim \Gamma(\delta_2, 1)$ are independent **Gamma variables**, another construction of a Feller–Pareto (FP) variable is^[7]

$$W = \mu + \sigma \left(\frac{U_1}{U_2} \right)^\gamma$$

and we write $W \sim FP(\mu, \sigma, \gamma, \delta_1, \delta_2)$. Special cases of the Feller–Pareto distribution are

$$FP(\sigma, \sigma, 1, 1, \alpha) = P(I)(\sigma, \alpha)$$

$$FP(\mu, \sigma, 1, 1, \alpha) = P(II)(\mu, \sigma, \alpha)$$

$$FP(\mu, \sigma, \gamma, 1, 1) = P(III)(\mu, \sigma, \gamma)$$

$$FP(\mu, \sigma, \gamma, 1, \alpha) = P(IV)(\mu, \sigma, \gamma, \alpha).$$

2.27.4 Applications

Pareto originally used this distribution to describe the **allocation of wealth** among individuals since it seemed to show rather well the way that a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. He also used it to describe distribution of income.^[8] This idea is sometimes expressed more simply as the **Pareto principle** or the “80-20 rule” which says that 20% of the population controls 80% of the wealth.^[9] However, the 80-20 rule corresponds to a particular value of α , and in fact, Pareto’s data on British income taxes in his *Cours d’économie politique* indicates that about 30% of the population had about 70% of the income. The **probability density function** (PDF) graph at the beginning of this article shows that the “probability” or fraction of the population that owns a small amount of wealth per person is rather high, and then decreases steadily as wealth increases. (Note that the Pareto distribution is not realistic for wealth for the lower end. In fact, **net worth** may even be negative.) This distribution is not limited to describing wealth or income, but to many situations in which an equilibrium is found in the distribution of the “small” to the “large”. The following examples are sometimes seen as approximately Pareto-distributed:

- The sizes of human settlements (few cities, many hamlets/villages)^[10]
- File size distribution of Internet traffic which uses the TCP protocol (many smaller files, few larger ones)^[10]
- **Hard disk drive** error rates^[11]
- Clusters of **Bose–Einstein condensate** near **absolute zero**^[12]
- The values of **oil reserves** in oil fields (a few **large fields**, many **small fields**)^[10]
- The length distribution in jobs assigned supercomputers (a few large ones, many small ones)
- The standardized price returns on individual stocks ^[10]

- Sizes of sand particles ^[10]
- Sizes of meteorites
- Numbers of species per genus (There is subjectivity involved: The tendency to divide a genus into two or more increases with the number of species in it)
- Areas burnt in forest fires
- Severity of large **casualty** losses for certain lines of business such as general liability, commercial auto, and workers compensation.^{[13][14]}
- In **hydrology** the Pareto distribution is applied to extreme events such as annually maximum one-day rainfalls and river discharges. The blue picture illustrates an example of fitting the Pareto distribution to ranked annually maximum one-day rainfalls showing also the 90% **confidence belt** based on the **binomial distribution**. The rainfall data are represented by **plotting positions** as part of the **cumulative frequency analysis**.

2.27.5 Relation to other distributions

Relation to the exponential distribution

The Pareto distribution is related to the **exponential distribution** as follows. If X is Pareto-distributed with minimum x_m and index α , then

$$Y = \log \left(\frac{X}{x_m} \right)$$

is exponentially distributed with rate parameter α . Equivalently, if Y is exponentially distributed with rate α , then

$$x_m e^Y$$

is Pareto-distributed with minimum x_m and index α .

This can be shown using the standard change of variable techniques:

$$\Pr(Y < y) = \Pr\left(\log\left(\frac{X}{x_m}\right) < y\right) = \Pr(X < x_m e^y) = 1 - \left(\frac{x_m}{x_m e^y}\right)^\alpha = 1 - e^{-\alpha y}.$$

The last expression is the cumulative distribution function of an exponential distribution with rate α .

Relation to the log-normal distribution

Note that the Pareto distribution and **log-normal distribution** are alternative distributions for describing the same types of quantities. One of the connections between the two is that they are both the distributions of the exponential of random variables distributed according to other common distributions, respectively the **exponential distribution** and **normal distribution**.

Relation to the generalized Pareto distribution

The Pareto distribution is a special case of the **generalized Pareto distribution**, which is a family of distributions of similar form, but containing an extra parameter in such a way that the support of the distribution is either bounded below (at a variable point), or bounded both above and below (where both are variable), with the **Lomax distribution** as a special case. This family also contains both the unshifted and shifted **exponential distributions**.

The Pareto distribution with scale x_m and shape α is equivalent to the generalized Pareto distribution with location $\mu = x_m$, scale $\sigma = x_m/\alpha$ and shape $\xi = 1/\alpha$. Vice versa one can get the Pareto distribution from the GPD by $x_m = \sigma/\xi$ and $\alpha = 1/\xi$.

Relation to Zipf's law

Pareto distributions are continuous probability distributions. **Zipf's law**, also sometimes called the **zeta distribution**, may be thought of as a discrete counterpart of the Pareto distribution.

Relation to the "Pareto principle"

The "**80-20 law**", according to which 20% of all people receive 80% of all income, and 20% of the most affluent 20% receive 80% of that 80%, and so on, holds precisely when the Pareto index is $\alpha = \log_4(5) = \log(5)/\log(4)$, approximately 1.161. This result can be derived from the **Lorenz curve** formula given below. Moreover, the following have been shown^[15] to be mathematically equivalent:

- Income is distributed according to a Pareto distribution with index $\alpha > 1$.
- There is some number $0 \leq p \leq 1/2$ such that $100p$ % of all people receive $100(1-p)$ % of all income, and similarly for every real (not necessarily integer) $n > 0$, $100p^n$ % of all people receive $100(1-p)^n$ percentage of all income.

This does not apply only to income, but also to wealth, or to anything else that can be modeled by this distribution.

This excludes Pareto distributions in which $0 < \alpha \leq 1$, which, as noted above, have infinite expected value, and so cannot reasonably model income distribution.

2.27.6 Lorenz curve and Gini coefficient

The **Lorenz curve** is often used to characterize income and wealth distributions. For any distribution, the Lorenz curve $L(F)$ is written in terms of the PDF f or the CDF F as

$$L(F) = \frac{\int_{x_m}^{x(F)} x f(x) dx}{\int_{x_m}^{\infty} x f(x) dx} = \frac{\int_0^F x(F') dF'}{\int_0^1 x(F') dF'}$$

where $x(F)$ is the inverse of the CDF. For the Pareto distribution,

$$x(F) = \frac{x_m}{(1 - F)^{\frac{1}{\alpha}}}$$

and the Lorenz curve is calculated to be

$$L(F) = 1 - (1 - F)^{1 - \frac{1}{\alpha}},$$

Although the numerator and denominator in the expression for $L(F)$ diverge for $0 \leq \alpha < 1$, their ratio does not, yielding $L=0$ in these cases, which yields a Gini coefficient of unity. Examples of the Lorenz curve for a number of Pareto distributions are shown in the graph on the right.

The **Gini coefficient** is a measure of the deviation of the Lorenz curve from the equidistribution line which is a line connecting $[0, 0]$ and $[1, 1]$, which is shown in black ($\alpha = \infty$) in the Lorenz plot on the right. Specifically, the Gini coefficient is twice the area between the Lorenz curve and the equidistribution line. The Gini coefficient for the Pareto distribution is then calculated (for $\alpha \geq 1$) to be

$$G = 1 - 2 \left(\int_0^1 L(F) dF \right) = \frac{1}{2\alpha - 1}$$

(see Aaberge 2005).

2.27.7 Parameter estimation

The **likelihood function** for the Pareto distribution parameters α and x_m , given a **sample** $x = (x_1, x_2, \dots, x_n)$, is

$$L(\alpha, x_m) = \prod_{i=1}^n \alpha \frac{x_m^\alpha}{x_i^{\alpha+1}} = \alpha^n x_m^{n\alpha} \prod_{i=1}^n \frac{1}{x_i^{\alpha+1}}.$$

Therefore, the logarithmic likelihood function is

$$\ell(\alpha, x_m) = n \ln \alpha + n\alpha \ln x_m - (\alpha + 1) \sum_{i=1}^n \ln x_i.$$

It can be seen that $\ell(\alpha, x_m)$ is monotonically increasing with x_m , that is, the greater the value of x_m , the greater the value of the likelihood function. Hence, since $x \geq x_m$, we conclude that

$$\hat{x}_m = \min_i x_i.$$

To find the **estimator** for α , we compute the corresponding partial derivative and determine where it is zero:

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + n \ln x_m - \sum_{i=1}^n \ln x_i = 0.$$

Thus the **maximum likelihood** estimator for α is:

$$\hat{\alpha} = \frac{n}{\sum_i (\ln x_i - \ln \hat{x}_m)}.$$

The expected statistical error is:^[16]

$$\sigma = \frac{\hat{\alpha}}{\sqrt{n}}.$$

Malik (1970)^[17] gives the exact joint distribution of $(\hat{x}_m, \hat{\alpha})$. In particular, $\hat{x}_m$ and $\hat{\alpha}$ are **independent** and $\hat{x}_m$ is Pareto with scale parameter x_m and shape parameter $n\alpha$, whereas $\hat{\alpha}$ has an **Inverse-gamma distribution** with shape and scale parameters $n-1$ and $n\alpha$, respectively.

2.27.8 Graphical representation

The characteristic curved '**Long Tail**' distribution when plotted on a linear scale, masks the underlying simplicity of the function when plotted on a **log-log graph**, which then takes the form of a straight line with negative gradient: It follows from the formula for the probability density function that for $x \geq x_m$,

$$\log f_X(x) = \log \left(\alpha \frac{x_m^\alpha}{x^{\alpha+1}} \right) = \log(\alpha x_m^\alpha) - (\alpha + 1) \log x.$$

Since α is positive, the gradient $-(\alpha+1)$ is negative.

2.27.9 Random sample generation

Random samples can be generated using **inverse transform sampling**. Given a random variate U drawn from the **uniform distribution** on the unit interval $(0, 1]$, the variate T given by

$$T = \frac{x_m}{U^{\frac{1}{\alpha}}}$$

is Pareto-distributed.^[18] If U is uniformly distributed on $[0, 1)$, it can be exchanged with $(1 - U)$.

2.27.10 Variants

Bounded Pareto distribution

See also: **Truncated distribution**

The bounded (or truncated) Pareto distribution has three parameters α , L and H . As in the standard Pareto distribution α determines the shape. L denotes the minimal value, and H denotes the maximal value. (The variance in the table on the right should be interpreted as the second moment).

The probability density function is

$$\frac{\alpha L^\alpha x^{-\alpha-1}}{1 - \left(\frac{L}{H}\right)^\alpha}$$

where $L \leq x \leq H$, and $\alpha > 0$.

Generating bounded Pareto random variables If U is uniformly distributed on $(0, 1)$, then applying inverse-transform method ^[19]

$$U = \frac{1 - L^\alpha x^{-\alpha}}{1 - \left(\frac{L}{H}\right)^\alpha}$$

$$x = \left(-\frac{UH^\alpha - UL^\alpha - H^\alpha}{H^\alpha L^\alpha} \right)^{-\frac{1}{\alpha}}$$

is a bounded Pareto-distributed.

Symmetric Pareto distribution

The symmetric Pareto distribution can be defined by the probability density function:^[20]

$$f(x; \alpha, x_m) = \begin{cases} \frac{1}{2} \alpha x_m^\alpha |x|^{-\alpha-1} & |x| > x_m \\ 0 & \text{otherwise.} \end{cases}$$

It has a similar shape to a Pareto distribution for $x > x_m$ and is mirror symmetric about the vertical axis.

2.27.11 See also

- Bradford's law
- Pareto analysis
- Pareto efficiency
- Pareto interpolation
- Power law probability distributions
- Traffic generation model

2.27.12 Notes

- [1] Barry C. Arnold (1983). *Pareto Distributions*. International Co-operative Publishing House. ISBN 0-89974-012-X.
- [2] Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions Vol 1. Wiley Series in Probability and Statistics.
- [3] Johnson, Kotz, and Balakrishnan (1994), (20.4).
- [4] Christian Kleiber and Samuel Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley. ISBN 0-471-15064-9.

- [5] Feller, W. (1971). *An Introduction to Probability Theory and its Applications II* (2nd ed.). New York: Wiley. p. 50. “The densities (4.3) are sometimes called after the economist *Pareto*. It was thought (rather naïvely from a modern statistical standpoint) that income distributions should have a tail with a density $\sim Ax^{-\alpha}$ as $x \rightarrow \infty$.”
- [6] Lomax, K. S. (1954). Business failures. Another example of the analysis of failure data. *Journal of the American Statistical Association*, 49, 847–852.
- [7] Chotikapanich, Duangkamon. “Chapter 7: Pareto and Generalized Pareto Distributions”. *Modeling Income Distributions and Lorenz Curves*. pp. 121–122.
- [8] Pareto, Vilfredo, *Cours d'Économie Politique: Nouvelle édition par G.-H. Bousquet et G. Busino*, Librairie Droz, Geneva, 1964, pages 299–345.
- [9] For a two-quantile population, where approximately 18% of the population owns 82% of the wealth, the *Theil index* takes the value 1.
- [10] Reed, William J.; et al. (2004.). “The Double Pareto-Lognormal Distribution – A New Parametric Model for Size Distributions”. *Communications in Statistics – Theory and Methods* **33** (8): 1733–1753. doi:10.1081/sta-120037438. CiteSeerX: 10.1.1.70.4555. Check date values in: `|date=` (help)
- [11] Schroeder, Bianca; Damouras, Sotirios; Gill, Phillipa (2010-02-24). “Understanding latent sector error and how to protect against them” (PDF). *8th Usenix Conference on File and Storage Technologies (FAST 2010)*. Retrieved 2010-09-10. We experimented with 5 different distributions (Geometric, Weibull, Rayleigh, Pareto, and Lognormal), that are commonly used in the context of system reliability, and evaluated their fit through the total squared differences between the actual and hypothesized frequencies (χ^2 statistic). We found consistently across all models that the geometric distribution is a poor fit, while the Pareto distribution provides the best fit.
- [12] Yuji Ijiri; Simon, Herbert A. (May 1975). “Some Distributions Associated with Bose–Einstein Statistics”. *Proc. Nat. Acad. Sci. USA* **72** (5): 1654–1657. PMC 432601. PMID 16578724. Retrieved 24 January 2013.
- [13] Kleiber and Kotz (2003): page 94.
- [14] Seal, H. (1980). “Survival probabilities based on Pareto claim distributions”. *ASTIN Bulletin* **11**: 61–71.
- [15] Hardy, Michael (2010). “Pareto’s Law”. *Mathematical Intelligencer* **32** (3): 38–43. doi:10.1007/s00283-010-9159-2.
- [16] M. E. J. Newman (2005). “Power laws, Pareto distributions and Zipf’s law”. *Contemporary Physics* **46** (5): 323–351. arXiv:cond-mat/0412004. Bibcode:2005ConPh..46..323N. doi:10.1080/00107510500052444.
- [17] H. J. Malik (1970). “Estimation of the Parameters of the Pareto Distribution”. *Metrika* **15**.
- [18] Tanizaki, Hisashi (2004). *Computational Methods in Statistics and Econometrics*. CRC Press. p. 133.
- [19] <http://www.cs.bgu.ac.il/~{ }mps042/invtransnote.htm>
- [20] Grabchak, M. & Samorodnitsky, D. “Do Financial Returns Have Finite or Infinite Variance? A Paradox and an Explanation” (PDF). pp. 7–8.

2.27.13 References

- M. O. Lorenz (1905). “Methods of measuring the concentration of wealth”. *Publications of the American Statistical Association* **9** (70): 209–219. Bibcode:1905PAmSA...9..209L. doi:10.2307/2276207.
- Pareto V (1965) “La Courbe de la Repartition de la Richesse” (Originally published in 1896). In: Busino G, editor. *Oevres Completes de Vilfredo Pareto*. Geneva: Librairie Droz. pp. 1–5.
- Pareto, V. (1895). La legge della domanda. *Giornale degli Economisti*, 10, 59–68. English translation in *Rivista di Politica Economica*, 87 (1997), 691–700.
- Pareto, V. (1897). *Cours d'économie politique*. Lausanne: Ed. Rouge.

2.27.14 External links

- Gini's Nuclear Family / Rolf Aabergé. – In: International Conference to Honor Two Eminent Social Scientists, May, 2005 – PDF
- Hazewinkel, Michiel, ed. (2001), “Pareto distribution”, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- `syntraf1.c` is a C program to generate synthetic packet traffic with bounded Pareto burst size and exponential interburst time.
- “Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes” /Mark E. Crovella and Azer Bestavros
- Weisstein, Eric W., “Pareto distribution”, *MathWorld*.

2.28 Pearson distribution

The **Pearson distribution** is a family of continuous probability distributions. It was first published by Karl Pearson in 1895 and subsequently extended by him in 1901 and 1916 in a series of articles on biostatistics.

2.28.1 History

The Pearson system was originally devised in an effort to model visibly **skewed** observations. It was well known at the time how to adjust a theoretical model to fit the first two **cumulants** or **moments** of observed data: Any **probability distribution** can be extended straightforwardly to form a **location-scale family**. Except in **pathological** cases, a location-scale family can be made to fit the observed **mean** (first cumulant) and **variance** (second cumulant) arbitrarily well. However, it was not known how to construct probability distributions in which the **skewness** (standardized third cumulant) and **kurtosis** (standardized fourth cumulant) could be adjusted equally freely. This need became apparent when trying to fit known theoretical models to observed data that exhibited skewness. Pearson's examples include survival data, which are usually asymmetric.

In his original paper, Pearson (1895, p. 360) identified four types of distributions (numbered I through IV) in addition to the **normal distribution** (which was originally known as type V). The classification depended on whether the distributions were **supported** on a bounded interval, on a half-line, or on the whole **real line**; and whether they were potentially skewed or necessarily symmetric. A second paper (Pearson 1901) fixed two omissions: it redefined the type V distribution (originally just the **normal distribution**, but now the **inverse-gamma distribution**) and introduced the type VI distribution. Together the first two papers cover the five main types of the Pearson system (I, III, IV, V, and VI). In a third paper, Pearson (1916) introduced further special cases and subtypes (VII through XII).

Rhind (1909, pp. 430–432) devised a simple way of visualizing the parameter space of the Pearson system, which was subsequently adopted by Pearson (1916, plate 1 and pp. 430ff., 448ff.). The Pearson types are characterized by two quantities, commonly referred to as β_1 and β_2 . The first is the square of the **skewness**: $\beta_1 = \gamma_1^2$ where γ_1 is the skewness, or third **standardized moment**. The second is the traditional **kurtosis**, or fourth standardized moment: $\beta_2 = \gamma_2 + 3$. (Modern treatments define kurtosis γ_2 in terms of cumulants instead of moments, so that for a normal distribution we have $\gamma_2 = 0$ and $\beta_2 = 3$. Here we follow the historical precedent and use β_2 .) The diagram on the right shows which Pearson type a given concrete distribution (identified by a point (β_1, β_2)) belongs to.

Many of the skewed and/or non-mesokurtic distributions familiar to us today were still unknown in the early 1890s. What is now known as the **beta distribution** had been used by Thomas Bayes as a **posterior distribution** of the parameter of a **Bernoulli distribution** in his 1763 work on inverse probability. The Beta distribution gained prominence due to its membership in Pearson's system and was known until the 1940s as the Pearson type I distribution. ^[1] (Pearson's type II distribution is a special case of type I, but is usually no longer singled out.) The **gamma distribution** originated from Pearson's work (Pearson 1893, p. 331; Pearson 1895, pp. 357, 360, 373–376) and was known as the Pearson type III distribution, before acquiring its modern name in the 1930s and 1940s. ^[2] Pearson's 1895 paper introduced the type IV

distribution, which contains **Student's t -distribution** as a special case, predating **William Sealy Gosset's** subsequent use by several years. His 1901 paper introduced the **inverse-gamma distribution** (type V) and the **beta prime distribution** (type VI).

2.28.2 Definition

A Pearson **density** p is defined to be any valid solution to the **differential equation** (cf. Pearson 1895, p. 381)

$$\frac{p'(x)}{p(x)} + \frac{a + x - \lambda}{b_2(x - \lambda)^2 + b_1(x - \lambda) + b_0} = 0. \quad (1)$$

with :

$$b_0 = \frac{4\beta_2 - 3\beta_1}{10\beta_2 - 12\beta_1 - 18} \mu_2,$$

$$a = b_1 = \sqrt{\mu_2} \sqrt{\beta_1} \frac{\beta_2 + 3}{10\beta_2 - 12\beta_1 - 18},$$

$$b_2 = \frac{2\beta_2 - 3\beta_1 - 6}{10\beta_2 - 12\beta_1 - 18}.$$

According to Ord,^[3] Pearson devised the underlying form of Equation (1) on the basis of, firstly, the formula for the derivative of the logarithm of the density function of the **normal distribution** (which gives a linear function) and, secondly, from a recurrence relation for values in the **probability mass function** of the **hypergeometric distribution** (which yields the linear-divided-by-quadratic structure).

In Equation (1), the parameter a determines a **stationary point**, and hence under some conditions a **mode** of the distribution, since

$$p'(\lambda - a) = 0$$

follows directly from the differential equation.

Since we are confronted with a **first order linear differential equation with variable coefficients**, its solution is straightforward:

$$p(x) \propto \exp \left(- \int \frac{x - a}{b_2 x^2 + b_1 x + b_0} dx \right).$$

The integral in this solution simplifies considerably when certain special cases of the integrand are considered. Pearson (1895, p. 367) distinguished two main cases, determined by the sign of the **discriminant** (and hence the number of real roots) of the **quadratic function**

$$f(x) = b_2 x^2 + b_1 x + b_0. \quad (2)$$

2.28.3 Particular types of distribution

Case 1, negative discriminant: The Pearson type IV distribution

If the discriminant of the quadratic function (2) is negative ($b_1^2 - 4b_2b_0 < 0$), it has no real roots. Then define

$$y = x + \frac{b_1}{2b_2} \text{ and}$$

$$\alpha = \frac{\sqrt{4b_2b_0 - b_1^2}}{2b_2}.$$

Observe that α is a well-defined real number and $\alpha \neq 0$, because by assumption $4b_2b_0 - b_1^2 > 0$ and therefore $b_2 \neq 0$. Applying these substitutions, the quadratic function (2) is transformed into

$$f(x) = b_2 (y^2 + \alpha^2).$$

The absence of real roots is obvious from this formulation, because α^2 is necessarily positive.

We now express the solution to the differential equation (1) as a function of y :

$$p(y) \propto \exp \left(-\frac{1}{b_2} \int \frac{y - \frac{b_1}{2b_2} - a}{y^2 + \alpha^2} dy \right).$$

Pearson (1895, p. 362) called this the “trigonometrical case”, because the integral

$$\int \frac{y - \frac{2b_2a+b_1}{2b_2}}{y^2 + \alpha^2} dy = \frac{1}{2} \ln(y^2 + \alpha^2) - \frac{2b_2a + b_1}{2b_2\alpha} \arctan \left(\frac{y}{\alpha} \right) + C_0$$

involves the **inverse trigonometric** arctan function. Then

$$p(y) \propto \exp \left[-\frac{1}{2b_2} \ln \left(1 + \frac{y^2}{\alpha^2} \right) - \frac{\ln \alpha}{b_2} + \frac{2b_2a + b_1}{2b_2^2\alpha} \arctan \left(\frac{y}{\alpha} \right) + C_1 \right]$$

Finally, let

$$m = \frac{1}{2b_2} \text{ and}$$

$$\nu = -\frac{2b_2a + b_1}{2b_2^2\alpha}$$

Applying these substitutions, we obtain the parametric function:

$$p(y) \propto \left[1 + \frac{y^2}{\alpha^2} \right]^{-m} \exp \left[-\nu \arctan \left(\frac{y}{\alpha} \right) \right]$$

This unnormalized density has **support** on the entire **real line**. It depends on a **scale parameter** $\alpha > 0$ and **shape parameters** $m > 1/2$ and ν . One parameter was lost when we chose to find the solution to the differential equation (1) as a function of y rather than x . We therefore reintroduce a fourth parameter, namely the **location parameter** λ . We have thus derived the density of the **Pearson type IV distribution**:

$$p(x) = \frac{\left| \frac{\Gamma(m + \frac{\nu}{2}i)}{\Gamma(m)} \right|^2}{\alpha B(m - \frac{1}{2}, \frac{1}{2})} \left[1 + \left(\frac{x - \lambda}{\alpha} \right)^2 \right]^{-m} \exp \left[-\nu \arctan \left(\frac{x - \lambda}{\alpha} \right) \right].$$

The normalizing constant involves the complex Gamma function (Γ) and the Beta function (B).

The Pearson type VII distribution The shape parameter ν of the Pearson type IV distribution controls its **skewness**. If we fix its value at zero, we obtain a symmetric three-parameter family. This special case is known as the **Pearson type VII distribution** (cf. Pearson 1916, p. 450). Its density is

$$p(x) = \frac{1}{\alpha B(m - \frac{1}{2}, \frac{1}{2})} \left[1 + \left(\frac{x - \lambda}{\alpha} \right)^2 \right]^{-m},$$

where B is the **Beta function**.

An alternative parameterization (and slight specialization) of the type VII distribution is obtained by letting

$$\alpha = \sigma \sqrt{2m - 3},$$

which requires $m > 3/2$. This entails a minor loss of generality but ensures that the **variance** of the distribution exists and is equal to σ^2 . Now the parameter m only controls the **kurtosis** of the distribution. If m approaches infinity as λ and σ are held constant, the **normal distribution** arises as a special case:

$$\begin{aligned} & \lim_{m \rightarrow \infty} \frac{1}{\sigma \sqrt{2m - 3} B(m - \frac{1}{2}, \frac{1}{2})} \left[1 + \left(\frac{x - \lambda}{\sigma \sqrt{2m - 3}} \right)^2 \right]^{-m} \\ &= \frac{1}{\sigma \sqrt{2} \Gamma(\frac{1}{2})} \times \lim_{m \rightarrow \infty} \frac{\Gamma(m)}{\Gamma(m - \frac{1}{2}) \sqrt{m - \frac{3}{2}}} \times \lim_{m \rightarrow \infty} \left[1 + \frac{\left(\frac{x - \lambda}{\sigma} \right)^2}{2m - 3} \right]^{-m} \\ &= \frac{1}{\sigma \sqrt{2} \pi} \times 1 \times \exp \left[-\frac{1}{2} \left(\frac{x - \lambda}{\sigma} \right)^2 \right] \end{aligned}$$

This is the density of a normal distribution with mean λ and standard deviation σ .

It is convenient to require that $m > 5/2$ and to let

$$m = \frac{5}{2} + \frac{3}{\gamma_2}.$$

This is another specialization, and it guarantees that the first four moments of the distribution exist. More specifically, the Pearson type VII distribution parameterized in terms of $(\lambda, \sigma, \gamma_2)$ has a mean of λ , **standard deviation** of σ , **skewness** of zero, and **excess kurtosis** of γ_2 .

Student's t -distribution The Pearson type VII distribution is equivalent to the non-standardized **Student's t -distribution** with parameters $\nu > 0$, μ , σ^2 by applying the following substitutions to its original parameterization:

$$\begin{aligned} \lambda &= \mu, \\ \alpha &= \sqrt{\nu \sigma^2}, \text{ and} \\ m &= \frac{\nu + 1}{2}, \end{aligned}$$

Observe that the constraint $m > 1/2$ is satisfied.

The resulting density is

$$p(x|\mu, \sigma^2, \nu) = \frac{1}{\sqrt{\nu \sigma^2} B(\frac{\nu}{2}, \frac{1}{2})} \left(1 + \frac{1}{\nu} \frac{(x - \mu)^2}{\sigma^2} \right)^{-\frac{\nu+1}{2}},$$

which is easily recognized as the density of a Student's t -distribution.

Note also that this implies that the Pearson type VII distribution subsumes the standard **Student's t -distribution** and also the standard **Cauchy distribution**. In particular, the standard Student's t -distribution arises as a subcase, when $\mu = 0$ and $\sigma^2 = 1$, equivalent to the following substitutions:

$$\begin{aligned}\lambda &= 0, \\ \alpha &= \sqrt{\nu}, \text{ and} \\ m &= \frac{\nu+1}{2},\end{aligned}$$

The density of this restricted one-parameter family is a standard Student's t :

$$p(x) = \frac{1}{\sqrt{\nu} B\left(\frac{\nu}{2}, \frac{1}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

Case 2, non-negative discriminant

If the quadratic function (2) has a non-negative discriminant ($b_1^2 - 4b_2b_0 \geq 0$), it has real roots a_1 and a_2 (not necessarily distinct):

$$a_1 = \frac{-b_1 - \sqrt{b_1^2 - 4b_2b_0}}{2b_2},$$

$$a_2 = \frac{-b_1 + \sqrt{b_1^2 - 4b_2b_0}}{2b_2},$$

In the presence of real roots the quadratic function (2) can be written as

$$f(x) = b_2 (x - a_1)(x - a_2),$$

and the solution to the differential equation is therefore

$$p(x) \propto \exp\left(-\frac{1}{b_2} \int \frac{x - a}{(x - a_1)(x - a_2)} dx\right).$$

Pearson (1895, p. 362) called this the “logarithmic case”, because the integral

$$\int \frac{x - a}{(x - a_1)(x - a_2)} dx = \frac{(a_1 - a) \ln(x - a_1) - (a_2 - a) \ln(x - a_2)}{a_1 - a_2} + C$$

involves only the **logarithm** function, and not the arctan function as in the previous case.

Using the substitution

$$\nu = \frac{1}{b_2 (a_1 - a_2)}$$

we obtain the following solution to the differential equation (1):

$$p(x) \propto (x - a_1)^{-\nu(a_1-a)} (x - a_2)^{\nu(a_2-a)}.$$

Since this density is only known up to a hidden constant of proportionality, that constant can be changed and the density written as follows:

$$p(x) \propto \left(1 - \frac{x}{a_1}\right)^{-\nu(a_1-a)} \left(1 - \frac{x}{a_2}\right)^{\nu(a_2-a)}$$

The Pearson type I distribution The **Pearson type I distribution** (a generalization of the **beta distribution**) arises when the roots of the quadratic equation (2) are of opposite sign, that is, $a_1 < 0 < a_2$. Then the solution p is supported on the interval (a_1, a_2) . Apply the substitution

$$x = a_1 + y(a_2 - a_1) \quad \text{where } 0 < y < 1,$$

which yields a solution in terms of y that is supported on the interval $(0, 1)$:

$$p(y) \propto \left(\frac{a_1 - a_2}{a_1} y\right)^{(-a_1+a)\nu} \left(\frac{a_2 - a_1}{a_2} (1 - y)\right)^{(a_2-a)\nu}.$$

One may define:

$$m_1 = \frac{a - a_1}{b_2(a_1 - a_2)}$$

$$m_2 = \frac{a - a_2}{b_2(a_2 - a_1)}$$

Regrouping constants and parameters, this simplifies to:

$$p(y) \propto y^{m_1} (1 - y)^{m_2},$$

Thus $\frac{x-\lambda-a_1}{a_2-a_1}$ follows a $B(m_1 + 1, m_2 + 1)$ with $\lambda = \mu_1 - (a_2 - a_1) \frac{m_1+1}{m_1+m_2+2} - a_1$

It turns out that $m_1, m_2 > -1$ is necessary and sufficient for p to be a proper probability density function.

The Pearson type II distribution The **Pearson type II distribution** is a special case of the Pearson type I family restricted to symmetric distributions.

For the Pearson Type II Curve,^[4]

$$y = y_0 \left(1 - \frac{x^2}{a^2}\right)^m$$

where

$$x = \sum d^2/2 - (n^3 - n)/12$$

the ordinate, y , is the frequency of $\sum d^2$. The Pearson Type II Curve is used in computing the table of significant correlation coefficients for **Spearman's rank correlation coefficient** when the number of items in a series is less than 100 (or 30, depending on some sources). After that, the distribution mimics a standard **Student's t-distribution**. For the table of values, certain values are used as the constants in the previous equation:

$$m = \frac{5\beta_2 - 9}{2(3 - \beta_2)}$$

$$a^2 = \frac{2\mu_2\beta_2}{3 - \beta_2}$$

$$y_0 = \frac{N[\Gamma(2m + 2)]}{a[2^{2m+1}]\Gamma(m + 1)}$$

The moments of x used are

$$\mu_2 = (n - 1)[(n^2 + n)/12]^2$$

$$\beta_2 = \frac{3(25n^4 - 13n^3 - 73n^2 + 37n + 72)}{25n(n + 1)^2(n - 1)}$$

The Pearson type III distribution

$$\lambda = \mu_1 + \frac{b_0}{b_1} - (m + 1)b_1$$

$$b_0 + b_1(x - \lambda) \text{ is Gamma}(m + 1, b_1^2)$$

The Pearson type III distribution is a **gamma distribution** or **chi-squared distribution**.

The Pearson type V distribution Defining new parameters:

$$C_1 = \frac{b_1}{2b_2}$$

$$\lambda = \mu_1 - \frac{a - C_1}{1 - 2b_2}$$

$$x - \lambda \text{ follows an InverseGamma}(\frac{1}{b_2} - 1, \frac{a - C_1}{b_2})$$

The Pearson type V distribution is an **inverse-gamma distribution**.

The Pearson type VI distribution

$$\lambda = \mu_1 + (a_2 - a_1) \frac{m_2 + 1}{m_2 + m_1 + 2} - a_2$$

$$\frac{x - \lambda - a_2}{a_2 - a_1} \text{ follows a } \beta'(m_2 + 1, -m_2 - m_1 - 1)$$

The Pearson type VI distribution is a **beta prime distribution** or **F-distribution**.

2.28.4 Relation to other distributions

The Pearson family subsumes the following distributions, among others:

- beta distribution (type I)

- beta prime distribution (type VI)
- Cauchy distribution (type IV)
- chi-squared distribution (type III)
- continuous uniform distribution (limit of type I)
- exponential distribution (type III)
- gamma distribution (type III)
- F -distribution (type VI)
- inverse-chi-squared distribution (type V)
- inverse-gamma distribution (type V)
- normal distribution (limit of type I, III, IV, V, or VI)
- Student's t -distribution (type VII, which is the non-skewed subtype of type IV)

2.28.5 Applications

These models are used in financial markets, given their ability to be parametrised in a way that has intuitive meaning for market traders. A number of models are in current use that capture the stochastic nature of the volatility of rates, stocks etc. and this family of distributions may prove to be one of the more important.

In the United States, the Log-Pearson III is the default distribution for flood frequency analysis.

2.28.6 Notes

- [1] Miller, Jeff; et al. (2006-07-09). "Beta distribution". Earliest Known Uses of Some of the Words of Mathematics. Retrieved December 9, 2006. External link in `|work=` (help)
- [2] Miller, Jeff; et al. (2006-12-07). "Gamma distribution". Earliest Known Uses of Some of the Words of Mathematics. Retrieved December 9, 2006. External link in `|work=` (help)
- [3] Ord J.K. (1972) p2
- [4] Ramsey, Philip H. (1989-09-01). "Critical Values for Spearman's Rank Order Correlation". Retrieved August 22, 2007.

2.28.7 Sources

Primary sources

- Pearson, Karl (1893). "Contributions to the mathematical theory of evolution [abstract]". *Proceedings of the Royal Society* **54** (326–330): 329–333. doi:10.1098/rspl.1893.0079. JSTOR 115538.
- Pearson, Karl (1895). "Contributions to the mathematical theory of evolution, II: Skew variation in homogeneous material". *Philosophical Transactions of the Royal Society* **186**: 343–414. Bibcode:1895RSPTA.186..343P. doi:10.1098/rsta.1895.0010. JSTOR 90649.
- Pearson, Karl (1901). "Mathematical contributions to the theory of evolution, X: Supplement to a memoir on skew variation". *Philosophical Transactions of the Royal Society A* **197** (287–299): 443–459. Bibcode:1901RSPTA.197..443P. doi:10.1098/rsta.1901.0023. JSTOR 90841.

- Pearson, Karl (1916). “Mathematical contributions to the theory of evolution, XIX: Second supplement to a memoir on skew variation”. *Philosophical Transactions of the Royal Society A* **216** (538–548): 429–457. Bibcode:1916RSPTA.216..429P. doi:10.1098/rsta.1916.0009. JSTOR 91092.
- Rhind, A. (July–October 1909). “Tables to facilitate the computation of the probable errors of the chief constants of skew frequency distributions”. *Biometrika* **7** (1/2): 127–147. doi:10.1093/biomet/7.1-2.127. JSTOR 2345367.

Secondary sources

- Milton Abramowitz and Irene A. Stegun (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards.
- Eric W. Weisstein et al. Pearson Type III Distribution. From MathWorld.

References

- Elderton, Sir W.P, Johnson, N.L. (1969) *Systems of Frequency Curves*. Cambridge University Press.
- Ord J.K. (1972) *Families of Frequency Distributions*. Griffin, London.

2.29 Phase-type distribution

A **phase-type distribution** is a probability distribution constructed by a convolution or mixture of exponential distributions.^[1] It results from a system of one or more inter-related Poisson processes occurring in sequence, or phases. The sequence in which each of the phases occur may itself be a stochastic process. The distribution can be represented by a random variable describing the time until absorption of a Markov process with one absorbing state. Each of the states of the Markov process represents one of the phases.

It has a discrete time equivalent the **discrete phase-type distribution**.

The set of phase-type distributions is dense in the field of all positive-valued distributions, that is, it can be used to approximate any positive-valued distribution.

2.29.1 Definition

Consider a continuous-time Markov process with $m + 1$ states, where $m \geq 1$, such that the states $1, \dots, m$ are transient states and state 0 is an absorbing state. Further, let the process have an initial probability of starting in any of the $m + 1$ phases given by the probability vector $(\alpha_0, \boldsymbol{\alpha})$ where α_0 is a scalar and $\boldsymbol{\alpha}$ is a $1 \times m$ vector.

The **continuous phase-type distribution** is the distribution of time from the above process's starting until absorption in the absorbing state.

This process can be written in the form of a transition rate matrix,

$$Q = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{S}^0 & S \end{bmatrix},$$

where S is an $m \times m$ matrix and $\mathbf{S}^0 = -S\mathbf{1}$. Here $\mathbf{1}$ represents an $m \times 1$ vector with every element being 1.

2.29.2 Characterization

The distribution of time X until the process reaches the absorbing state is said to be phase-type distributed and is denoted $\text{PH}(\alpha, S)$.

The distribution function of X is given by,

$$F(x) = 1 - \alpha \exp(Sx) \mathbf{1},$$

and the density function,

$$f(x) = \alpha \exp(Sx) S^0,$$

for all $x > 0$, where $\exp(\cdot)$ is the **matrix exponential**. It is usually assumed the probability of process starting in the absorbing state is zero (i.e. $\alpha_0 = 0$). The moments of the distribution function are given by

$$E[X^n] = (-1)^n n! \alpha S^{-n} \mathbf{1}.$$

2.29.3 Special cases

The following probability distributions are all considered special cases of a continuous phase-type distribution:

- **Degenerate distribution**, point mass at zero or the **empty phase-type distribution** - 0 phases.
- **Exponential distribution** - 1 phase.
- **Erlang distribution** - 2 or more identical phases in sequence.
- **Deterministic distribution** (or constant) - The limiting case of an Erlang distribution, as the number of phases become infinite, while the time in each state becomes zero.
- **Coxian distribution** - 2 or more (not necessarily identical) phases in sequence, with a probability of transitioning to the terminating/absorbing state after each phase.
- **Hyper-exponential distribution** (also called a mixture of exponential) - 2 or more non-identical phases, that each have a probability of occurring in a mutually exclusive, or parallel, manner. (Note: The exponential distribution is the degenerate situation when all the parallel phases are identical.)
- **Hypoexponential distribution** - 2 or more phases in sequence, can be non-identical or a mixture of identical and non-identical phases, generalises the Erlang.

As the phase-type distribution is dense in the field of all positive-valued distributions, we can represent any positive valued distribution. However, the phase-type is a light-tailed or platikurtic distribution. So the representation of heavy-tailed or leptokurtic distribution by phase type is an approximation, even if the precision of the approximation can be as good as we want.

2.29.4 Examples

In all the following examples it is assumed that there is no probability mass at zero, that is $\alpha_0 = 0$.

Exponential distribution

The simplest non-trivial example of a phase-type distribution is the exponential distribution of parameter λ . The parameter of the phase-type distribution are : $S = -\lambda$ and $\alpha = 1$.

Hyper-exponential or mixture of exponential distribution

The mixture of exponential or **hyper-exponential distribution** with $\lambda_1, \lambda_2, \dots, \lambda_n > 0$ can be represented as a phase type distribution with

$$\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_n)$$

with $\sum_{i=1}^n \alpha_i = 1$ and

$$S = \begin{bmatrix} -\lambda_1 & 0 & 0 & 0 & 0 \\ 0 & -\lambda_2 & 0 & 0 & 0 \\ 0 & 0 & -\lambda_3 & 0 & 0 \\ 0 & 0 & 0 & -\lambda_4 & 0 \\ 0 & 0 & 0 & 0 & -\lambda_5 \end{bmatrix}.$$

This mixture of densities of exponential distributed random variables can be characterized through

$$f(x) = \sum_{i=1}^n \alpha_i \lambda_i e^{-\lambda_i x} = \sum_{i=1}^n \alpha_i f_{X_i}(x),$$

or its cumulative distribution function

$$F(x) = 1 - \sum_{i=1}^n \alpha_i e^{-\lambda_i x} = \sum_{i=1}^n \alpha_i F_{X_i}(x).$$

with $X_i \sim \text{Exp}(\lambda_i)$

Erlang distribution

The Erlang distribution has two parameters, the shape an integer $k > 0$ and the rate $\lambda > 0$. This is sometimes denoted $E(k, \lambda)$. The Erlang distribution can be written in the form of a phase-type distribution by making S a $k \times k$ matrix with diagonal elements $-\lambda$ and super-diagonal elements λ , with the probability of starting in state 1 equal to 1. For example $E(5, \lambda)$,

$$\alpha = (1, 0, 0, 0, 0),$$

and

$$S = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 \\ 0 & -\lambda & \lambda & 0 & 0 \\ 0 & 0 & -\lambda & \lambda & 0 \\ 0 & 0 & 0 & -\lambda & \lambda \\ 0 & 0 & 0 & 0 & -\lambda \end{bmatrix}.$$

For a given number of phases, the Erlang distribution is the phase type distribution with smallest coefficient of variation.^[2]

The **hypoexponential distribution** is a generalisation of the Erlang distribution by having different rates for each transition (the non-homogeneous case).

Mixture of Erlang distribution

The mixture of two Erlang distribution with parameter $E(3, \beta_1)$, $E(3, \beta_2)$ and (α_1, α_2) (such that $\alpha_1 + \alpha_2 = 1$ and for each i , $\alpha_i \geq 0$) can be represented as a phase type distribution with

$$\alpha = (\alpha_1, 0, 0, \alpha_2, 0, 0),$$

and

$$S = \begin{bmatrix} -\beta_1 & \beta_1 & 0 & 0 & 0 & 0 \\ 0 & -\beta_1 & \beta_1 & 0 & 0 & 0 \\ 0 & 0 & -\beta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta_2 & \beta_2 & 0 \\ 0 & 0 & 0 & 0 & -\beta_2 & \beta_2 \\ 0 & 0 & 0 & 0 & 0 & -\beta_2 \end{bmatrix}.$$

Coxian distribution

The **Coxian distribution** is a generalisation of the **hypoexponential distribution**. Instead of only being able to enter the absorbing state from state k it can be reached from any phase. The phase-type representation is given by,

$$S = \begin{bmatrix} -\lambda_1 & p_1\lambda_1 & 0 & \dots & 0 & 0 \\ 0 & -\lambda_2 & p_2\lambda_2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & -\lambda_{k-2} & p_{k-2}\lambda_{k-2} & 0 \\ 0 & 0 & \dots & 0 & -\lambda_{k-1} & p_{k-1}\lambda_{k-1} \\ 0 & 0 & \dots & 0 & 0 & -\lambda_k \end{bmatrix}$$

and

$$\alpha = (1, 0, \dots, 0),$$

where $0 < p_1, \dots, p_{k-1} \leq 1$. In the case where all $p_i = 1$ we have the hypoexponential distribution. The Coxian distribution is extremely important as any acyclic phase-type distribution has an equivalent Coxian representation.

The **generalised Coxian distribution** relaxes the condition that requires starting in the first phase.

2.29.5 Generating samples from phase-type distributed random variables

BuTools includes methods for generating samples from phase-type distributed random variables.^[3]

2.29.6 Approximating other distributions

Any distribution can be arbitrarily well approximated by a phase type distribution.^{[4][5]} In practice, however, approximations can be poor when the size of the approximating process is fixed. Approximating a deterministic distribution of time 1 with 10 phases, each of average length 0.1 will have variance 0.1 (because the Erlang distribution has smallest variance^[2]).

- *BuTools* a **MATLAB** and **Mathematica** script for fitting phase-type distributions to 3 specified moments
- *momentmatching* a **MATLAB** script to fit a minimal phase-type distribution to 3 specified moments^[6]

2.29.7 Fitting a phase type distribution to data

Methods to fit a phase type distribution to data can be classified as maximum likelihood methods or moment matching methods.^[7] Fitting a phase type distribution to **heavy-tailed distributions** has been shown to be practical in some situations.^[8]

- *PhFit* a C script for fitting discrete and continuous phase type distributions to data^[9]
- *EMpht* is a C script for fitting phase-type distributions to data or parametric distributions using an **expectation–maximization algorithm**.^[10]
- *HyperStar* was developed around the core idea of making phase-type fitting simple and user-friendly, in order to advance the use of phase-type distributions in a wide range of areas. It provides a graphical user interface and yields good fitting results with only little user interaction.^[11]
- *jPhase* is a Java library which can also compute metrics for queues using the fitted phase type distribution^[12]

2.29.8 See also

- Discrete phase-type distribution
- Continuous-time Markov process
- Exponential distribution
- Hyper-exponential distribution
- Queueing theory

2.29.9 References

- [1] Harchol-Balter, M. (2012). “Real-World Workloads: High Variability and Heavy Tails”. *Performance Modeling and Design of Computer Systems*. p. 347. doi:10.1017/CBO9781139226424.026. ISBN 9781139226424.
- [2] Aldous, David; Shepp, Larry (1987). “The least variable phase type distribution is erlang” (PDF). *Stochastic Models* **3** (3): 467. doi:10.1080/15326348708807067.
- [3] Horváth, G. B.; Reinecke, P.; Telek, M. S.; Wolter, K. (2012). “Efficient Generation of PH-Distributed Random Variates”. *Analytical and Stochastic Modeling Techniques and Applications*. Lecture Notes in Computer Science **7314**. p. 271. doi:10.1007/978-3-642-30782-9_19. ISBN 978-3-642-30781-2.
- [4] Bolch, Gunter; Greiner, Stefan; de Meer, Hermann; Trivedi, Kishor S. (1998). “Steady-State Solutions of Markov Chains”. *Queueing Networks and Markov Chains*. pp. 103–151. doi:10.1002/0471200581.ch3. ISBN 0471193666.
- [5] Cox, D. R. (2008). “A use of complex probabilities in the theory of stochastic processes”. *Mathematical Proceedings of the Cambridge Philosophical Society* **51** (2): 313. doi:10.1017/S0305004100030231.

- [6] Osogami, T.; Harchol-Balter, M. (2006). “Closed form solutions for mapping general distributions to quasi-minimal PH distributions”. *Performance Evaluation* **63** (6): 524. doi:10.1016/j.peva.2005.06.002.
 - [7] Lang, Andreas; Arthur, Jeffrey L. (1996). “Parameter approximation for Phase-Type distributions”. In Chakravarthy, S.; Alfa, Attahiru S. *Matrix Analytic methods in Stochastic Models*. CRC Press. ISBN 0824797663.
 - [8] Ramaswami, V.; Poole, D.; Ahn, S.; Byers, S.; Kaplan, A. (2005). “Ensuring Access to Emergency Services in the Presence of Long Internet Dial-Up Calls”. *Interfaces* **35** (5): 411. doi:10.1287/inte.1050.0155.
 - [9] Horváth, András S.; Telek, Miklós S. (2002). “PhFit: A General Phase-Type Fitting Tool”. *Computer Performance Evaluation: Modelling Techniques and Tools*. Lecture Notes in Computer Science **2324**. p. 82. doi:10.1007/3-540-46029-2_5. ISBN 978-3-540-43539-6.
 - [10] Asmussen, Søren; Nerman, Olle; Olsson, Marita (1996). “Fitting Phase-Type Distributions via the EM Algorithm”. *Scandinavian Journal of Statistics* **23** (4): 419–441. JSTOR 4616418.
 - [11] Reinecke, P.; Krauß, T.; Wolter, K. (2012). “Cluster-based fitting of phase-type distributions to empirical data”. *Computers & Mathematics with Applications* **64** (12): 3840. doi:10.1016/j.camwa.2012.03.016.
 - [12] Pérez, J. F.; Riaño, G. N. (2006). “jPhase: an object-oriented tool for modeling phase-type distributions”. *Proceeding from the 2006 workshop on Tools for solving structured Markov chains (SMCtools '06)* (PDF). doi:10.1145/1190366.1190370. ISBN 1595935061.
- M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach*, Chapter 2: Probability Distributions of Phase Type; Dover Publications Inc., 1981.
 - G. Latouche, V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modelling*, 1st edition. Chapter 2: PH Distributions; ASA SIAM, 1999.
 - C. A. O’Cinneide (1990). *Characterization of phase-type distributions*. *Communications in Statistics: Stochastic Models*, **6**(1), 1-57.
 - C. A. O’Cinneide (1999). *Phase-type distribution: open problems and a few properties*, *Communication in Statistic: Stochastic Models*, **15**(4), 731-757.

2.30 Rayleigh distribution

Not to be confused with [Rayleigh mixture distribution](#).

In probability theory and statistics, the **Rayleigh distribution** /ˈreɪli/ is a continuous probability distribution for positive-valued random variables.

A Rayleigh distribution is often observed when the overall magnitude of a vector is related to its directional components. One example where the Rayleigh distribution naturally arises is when wind velocity is analyzed into its orthogonal 2-dimensional vector components. Assuming that each component is uncorrelated, normally distributed with equal variance, and zero mean, then the overall wind speed (vector magnitude) will be characterized by a Rayleigh distribution. A second example of the distribution arises in the case of random complex numbers whose real and imaginary components are i.i.d. (independently and identically distributed) Gaussian with equal variance and zero mean. In that case, the absolute value of the complex number is Rayleigh-distributed.

The distribution is named after Lord Rayleigh.

2.30.1 Definition

The probability density function of the Rayleigh distribution is^[1]

$$f(x; \sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0,$$

where σ is the **scale parameter** of the distribution. The **cumulative distribution function** is^[1]

$$F(x; \sigma) = 1 - e^{-x^2/(2\sigma^2)}$$

for $x \in [0, \infty)$.

2.30.2 Relation to random vector lengths

Consider the two-dimensional vector $Y = (U, V)$ which has components that are Gaussian-distributed, centered at zero, and independent. Then $f_U(u; \sigma) = \frac{e^{-u^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$, and similarly for $f_V(v; \sigma)$.

Let x be the length of Y . It is distributed as

$$f(x; \sigma) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv e^{-u^2/2\sigma^2} e^{-v^2/2\sigma^2} \delta(x - \sqrt{u^2 + v^2}).$$

By transforming to the **polar coordinate system** one has

$$f(x; \sigma) = \frac{1}{2\pi\sigma^2} \int_0^{2\pi} d\phi \int_0^{\infty} dr \delta(r - x) r e^{-r^2/2\sigma^2} = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2},$$

which is the Rayleigh distribution. It is straightforward to generalize to vectors of dimension other than 2. There are also generalizations when the components have **unequal variance** or correlations.

2.30.3 Properties

The raw **moments** are given by:

$$\mu_k = \sigma^k 2^{\frac{k}{2}} \Gamma\left(1 + \frac{k}{2}\right)$$

where $\Gamma(z)$ is the **Gamma function**.

The **mean** and **variance** of a Rayleigh random variable may be expressed as:

$$\mu(X) = \sigma \sqrt{\frac{\pi}{2}} \approx 1.253\sigma$$

and

$$\text{var}(X) = \frac{4 - \pi}{2} \sigma^2 \approx 0.429\sigma^2$$

The mode is σ and the maximum pdf is

$$f_{\max} = f(\sigma; \sigma) = \frac{1}{\sigma} e^{-\frac{1}{2}} \approx \frac{1}{\sigma} 0.606$$

The **skewness** is given by:

$$\gamma_1 = \frac{2\sqrt{\pi}(\pi - 3)}{(4 - \pi)^{\frac{3}{2}}} \approx 0.631$$

The excess **kurtosis** is given by:

$$\gamma_2 = -\frac{6\pi^2 - 24\pi + 16}{(4 - \pi)^2} \approx 0.245$$

The **characteristic function** is given by:

$$\varphi(t) = 1 - \sigma t e^{-\frac{1}{2}\sigma^2 t^2} \sqrt{\frac{\pi}{2}} \left[\operatorname{erfi}\left(\frac{\sigma t}{\sqrt{2}}\right) - i \right]$$

where $\operatorname{erfi}(z)$ is the imaginary **error function**. The **moment generating function** is given by

$$M(t) = 1 + \sigma t e^{\frac{1}{2}\sigma^2 t^2} \sqrt{\frac{\pi}{2}} \left[\operatorname{erf}\left(\frac{\sigma t}{\sqrt{2}}\right) + 1 \right]$$

where $\operatorname{erf}(z)$ is the **error function**.

Differential entropy

The **differential entropy** is given by

$$H = 1 + \ln\left(\frac{\sigma}{\sqrt{2}}\right) + \frac{\gamma}{2}$$

where γ is the **Euler–Mascheroni constant**.

Differential equation

The pdf of the Rayleigh distribution is a solution of the following **differential equation**:

$$\left\{ \begin{array}{l} \sigma^2 x f'(x) + f(x) (x^2 - \sigma^2) = 0 \\ f(1) = \frac{\exp\left(-\frac{1}{2\sigma^2}\right)}{\sigma^2} \end{array} \right\}$$

2.30.4 Parameter estimation

Given a sample of N independent and identically distributed Rayleigh random variables x_i with parameter σ ,

$\hat{\sigma}^2 \approx \frac{1}{2N} \sum_{i=1}^N x_i^2$ is an unbiased maximum likelihood estimate.

$$\hat{\sigma} \approx \sqrt{\frac{1}{2N} \sum_{i=1}^N x_i^2}$$

$$\sigma = \hat{\sigma} \frac{\Gamma(N)\sqrt{N}}{\Gamma(N+\frac{1}{2})} = \hat{\sigma} \frac{4^N N!(N-1)!\sqrt{N}}{(2N)!\sqrt{\pi}} \quad [2]$$

Confidence intervals

To find the $(1 - \alpha)$ confidence interval, first find the two numbers χ_1^2, χ_2^2 where:

$$Pr(\chi^2(2N) \leq \chi_1^2) = \alpha/2, \quad Pr(\chi^2(2N) \leq \chi_2^2) = 1 - \alpha/2$$

then

$$\frac{Nx^2}{\chi_2^2} \leq \hat{\sigma}^2 \leq \frac{Nx^2}{\chi_1^2} \quad [3]$$

2.30.5 Generating random variates

Given a random variate U drawn from the uniform distribution in the interval $(0, 1)$, then the variate

$$X = \sigma \sqrt{-2 \ln(U)}$$

has a Rayleigh distribution with parameter σ . This is obtained by applying the inverse transform sampling-method.

2.30.6 Related distributions

- $R \sim \text{Rayleigh}(\sigma)$ is Rayleigh distributed if $R = \sqrt{X^2 + Y^2}$, where $X \sim N(0, \sigma^2)$ and $Y \sim N(0, \sigma^2)$ are independent normal random variables.^[4] (This gives motivation to the use of the symbol “sigma” in the above parameterization of the Rayleigh density.)
- The chi distribution with $\nu = 2$ is equivalent to Rayleigh Distribution with $\sigma = 1$. I.e., if $R \sim \text{Rayleigh}(1)$, then R^2 has a chi-squared distribution with parameter N , degrees of freedom, equal to two ($N = 2$)

$$[Q = R^2] \sim \chi^2(N).$$

- If $R \sim \text{Rayleigh}(\sigma)$, then $\sum_{i=1}^N R_i^2$ has a gamma distribution with parameters N and $2\sigma^2$

$$\left[Y = \sum_{i=1}^N R_i^2 \right] \sim \Gamma(N, 2\sigma^2).$$

- The Rice distribution is a generalization of the Rayleigh distribution.
- The Weibull distribution is a generalization of the Rayleigh distribution. In this instance, parameter σ is related to the Weibull scale parameter λ : $\lambda = \sigma\sqrt{2}$.
- The Maxwell–Boltzmann distribution describes the magnitude of a normal vector in three dimensions.
- If X has an exponential distribution $X \sim \text{Exponential}(\lambda)$, then $Y = \sqrt{2X\sigma^2\lambda} \sim \text{Rayleigh}(\sigma)$.

2.30.7 Applications

An application of the estimation of σ can be found in **magnetic resonance imaging** (MRI). As MRI images are recorded as **complex** images but most often viewed as magnitude images, the background data is Rayleigh distributed. Hence, the above formula can be used to estimate the noise variance in an MRI image from background data.^{[5] [6]}

2.30.8 Proof of correctness – Unequal variances

Click [show] to expand

We start with

$$f(x; \sigma) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv e^{-u^2/2\sigma_1^2} e^{-v^2/2\sigma_2^2} \delta(x - \sqrt{u^2 + v^2}).$$

as above, except with σ_1 and σ_2 distinct.

Let $a = u\sigma_2/\sigma_1$ so that $a/\sigma_2 = u/\sigma_1$, and differentiating we have:

$$\frac{\sigma_1}{\sigma_2} da = du$$

Substituting,

$$f(x; \sigma) = \frac{\sigma_1}{\sigma_2} \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} da \int_{-\infty}^{\infty} dv e^{-a^2/2\sigma_2^2} e^{-v^2/2\sigma_2^2} \delta\left(x - \sqrt{\left(\frac{\sigma_1}{\sigma_2}a\right)^2 + v^2}\right)$$

As before, we perform a polar coordinate transformation:^[7]

$$\begin{cases} a = r\cos(\phi) \\ v = r\sin(\phi) \\ da dv = r dr d\phi \end{cases}$$

Substituting,

$$f(x; \sigma) = \frac{\sigma_1}{\sigma_2} \frac{1}{2\pi\sigma_1\sigma_2} \int_0^{2\pi} d\phi \int_0^{\infty} r dr e^{-r^2/2\sigma_2^2} \delta\left(x - \sqrt{\left(\frac{\sigma_1}{\sigma_2}a\right)^2 + v^2}\right).$$

Simplifying,

$$f(x; \sigma) = \frac{1}{2\pi\sigma_2^2} \int_0^{2\pi} d\phi \int_0^{\infty} r dr e^{-r^2/2\sigma_2^2} \delta\left(x - \sqrt{\left(\frac{\sigma_1}{\sigma_2}a\right)^2 + v^2}\right).$$

See Hoyt distribution for more information.

2.30.9 See also

- Normal distribution
- Rayleigh fading
- Rayleigh mixture distribution
- Circular error probable

2.30.10 References

- [1] Papoulis, Athanasios; Pillai, S. (2001) *Probability, Random Variables and Stochastic Processes*. ISBN 0073660116, ISBN 9780073660110
- [2] Siddiqui, M. M. (1964) “Statistical inference for Rayleigh distributions”, *The Journal of Research of the National Bureau of Standards, Sec. D: Radio Science*, Vol. 68D, No. 9, p. 1007
- [3] Siddiqui, M. M. (1961) “Some Problems Connected With Rayleigh Distributions”, *The Journal of Research of the National Bureau of Standards, Sec. D: Radio Propagation*, Vol. 66D, No. 2, p. 169
- [4] Hogema, Jeroen (2005) “Shot group statistics”
- [5] Sijbers J., den Dekker A. J., Raman E. and Van Dyck D. (1999) “Parameter estimation from magnitude MR images”, *International Journal of Imaging Systems and Technology*, 10(2), 109–114
- [6] den Dekker A. J., Sijbers J., (2014) “Data distributions in magnetic resonance images: a review”, *Physica Medica*,
- [7] <http://physicspages.com/2012/12/24/coordinate-transformations-the-jacobian-determinant/>

2.31 Rayleigh mixture distribution

Not to be confused with Rayleigh scattering.

In probability theory and statistics a **Rayleigh mixture distribution** is a weighted mixture of multiple probability distributions where the weightings are equal to the weightings of a Rayleigh distribution.^[1] Since the probability density function for a (standard) Rayleigh distribution is given by^[2]

$$f(x; \sigma) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, \quad x \geq 0,$$

Rayleigh mixture distributions have probability density functions of the form

$$f(x; \sigma, n) = \int_0^\infty \frac{r e^{-r^2/2\sigma^2}}{\sigma^2} \tau(x, r; n) dr,$$

where $\tau(x, r; n)$ is a well-defined probability density function or sampling distribution.^[1]

The Rayleigh mixture distribution is one of many types of compound distributions in which the appearance of a value in a sample or population might be interpreted as a function of other underlying random variables. Mixture distributions are often used in mixture models, which are used to express probabilities of sub-populations within a larger population.

2.31.1 See also

- Mixture distribution
- List of probability distributions

2.31.2 References

- [1] Karim R., Hossain P., Begum S., and Hossain F., “Rayleigh Mixture Distribution”, Journal of Applied Mathematics, Vol. 2011, doi:10.1155/2011/238290 (2011).
- [2] Jackson J.L., “Properties of the Rayleigh Distribution”, Johns Hopkins University (1954).

2.32 Rice distribution

In probability theory, the **Rice distribution** or **Rician distribution** is the probability distribution of the magnitude of a circular bivariate normal random variable with potentially non-zero mean. It was named after **Stephen O. Rice**.

2.32.1 Characterization

The probability density function is

$$f(x \mid \nu, \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{(x^2 + \nu^2)}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right),$$

where $I_0(z)$ is the modified Bessel function of the first kind with order zero.

The characteristic function is:^{[1][2]}

$$\begin{aligned} \chi_X(t \mid \nu, \sigma) &= \exp\left(-\frac{\nu^2}{2\sigma^2}\right) \left[\Psi_2\left(1; 1, \frac{1}{2}; \frac{\nu^2}{2\sigma^2}, -\frac{1}{2}\sigma^2 t^2\right) \right. \\ &\quad \left. + i\sqrt{2}\sigma t \Psi_2\left(\frac{3}{2}; 1, \frac{3}{2}; \frac{\nu^2}{2\sigma^2}, -\frac{1}{2}\sigma^2 t^2\right) \right], \end{aligned}$$

where $\Psi_2(\alpha; \gamma, \gamma'; x, y)$ is one of Horn's confluent hypergeometric functions with two variables and convergent for all finite values of x and y . It is given by:^{[3][4]}

$$\Psi_2(\alpha; \gamma, \gamma'; x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{(\alpha)_{m+n}}{(\gamma)_m (\gamma')_n} \frac{x^m y^n}{m! n!},$$

where

$$(x)_n = x(x+1) \cdots (x+n-1) = \frac{\Gamma(x+n)}{\Gamma(x)}$$

is the rising factorial.

2.32.2 Properties

Moments

The first few raw moments are:

$$\mu_1' = \sigma \sqrt{\pi/2} L_{1/2}(-\nu^2/2\sigma^2)$$

$$\mu_2' = 2\sigma^2 + \nu^2$$

$$\mu_3' = 3\sigma^3 \sqrt{\pi/2} L_{3/2}(-\nu^2/2\sigma^2)$$

$$\mu_4' = 8\sigma^4 + 8\sigma^2\nu^2 + \nu^4$$

$$\mu_5' = 15\sigma^5 \sqrt{\pi/2} L_{5/2}(-\nu^2/2\sigma^2)$$

$$\mu_6' = 48\sigma^6 + 72\sigma^4\nu^2 + 18\sigma^2\nu^4 + \nu^6$$

and, in general, the raw moments are given by

$$\mu_k' = \sigma^k 2^{k/2} \Gamma(1+k/2) L_{k/2}(-\nu^2/2\sigma^2).$$

Here $L_q(x)$ denotes a **Laguerre polynomial**:

$$L_q(x) = L_q^{(0)}(x) = M(-q, 1, x) = {}_1F_1(-q; 1; x)$$

where $M(a, b, z) = {}_1F_1(a; b; z)$ is the **confluent hypergeometric function** of the first kind. When k is even, the raw moments become simple polynomials in σ and ν , as in the examples above.

For the case $q = 1/2$:

$$\begin{aligned} L_{1/2}(x) &= {}_1F_1\left(-\frac{1}{2}; 1; x\right) \\ &= e^{x/2} \left[(1-x) I_0\left(\frac{-x}{2}\right) - x I_1\left(\frac{-x}{2}\right) \right]. \end{aligned}$$

The second **central moment**, the **variance**, is

$$\mu_2 = 2\sigma^2 + \nu^2 - (\pi\sigma^2/2) L_{1/2}^2(-\nu^2/2\sigma^2).$$

Note that $L_{1/2}^2(\cdot)$ indicates the square of the Laguerre polynomial $L_{1/2}(\cdot)$, not the generalized Laguerre polynomial $L_{1/2}^{(2)}(\cdot)$.

Differential equation

The pdf of the Rice distribution is a solution of the following **differential equation**:

$$\left\{ \begin{array}{l} \sigma^4 x^2 f''(x) + (2\sigma^2 x^3 - \sigma^4 x) f'(x) + f(x) (\sigma^4 - \nu^2 x^2 + x^4) = 0 \\ f(1) = \frac{\exp\left(-\frac{\nu^2+1}{2\sigma^2}\right) I_0\left(\frac{\nu}{\sigma^2}\right)}{\sigma^2} \\ f'(1) = \frac{\exp\left(-\frac{\nu^2+1}{2\sigma^2}\right) ((\sigma^2-1) I_0\left(\frac{\nu}{\sigma^2}\right) + \nu I_1\left(\frac{\nu}{\sigma^2}\right))}{\sigma^4} \end{array} \right\}$$

2.32.3 Related distributions

- $R \sim \text{Rice}(\nu, \sigma)$ has a Rice distribution if $R = \sqrt{X^2 + Y^2}$ where $X \sim N(\nu \cos \theta, \sigma^2)$ and $Y \sim N(\nu \sin \theta, \sigma^2)$ are statistically independent normal random variables and θ is any real number.
- Another case where $R \sim \text{Rice}(\nu, \sigma)$ comes from the following steps:
 1. Generate P having a **Poisson distribution** with parameter (also mean, for a Poisson) $\lambda = \frac{\nu^2}{2\sigma^2}$.
 2. Generate X having a **chi-squared distribution** with $2P + 2$ degrees of freedom.

$$R = \sigma\sqrt{X}.$$

- If $R \sim \text{Rice}(\nu, 1)$ then R^2 has a **noncentral chi-squared distribution** with two degrees of freedom and noncentrality parameter ν^2 .
- If $R \sim \text{Rice}(\nu, 1)$ then R has a **noncentral chi distribution** with two degrees of freedom and noncentrality parameter ν .
- If $R \sim \text{Rice}(0, \sigma)$ then $R \sim \text{Rayleigh}(\sigma)$, i.e., for the special case of the Rice distribution given by $\nu = 0$, the distribution becomes the **Rayleigh distribution**, for which the variance is $\mu_2 = \frac{4-\pi}{2}\sigma^2$.
- If $R \sim \text{Rice}(0, \sigma)$ then R^2 has an **exponential distribution**.^[5]

2.32.4 Limiting cases

For large values of the argument, the Laguerre polynomial becomes^[6]

$$\lim_{x \rightarrow -\infty} L_\nu(x) = \frac{|x|^\nu}{\Gamma(1 + \nu)}.$$

It is seen that as ν becomes large or σ becomes small the mean becomes ν and the variance becomes σ^2 .

2.32.5 Parameter estimation (the Koay inversion technique)

There are three different methods for estimating the parameters of the Rice distribution, (1) **method of moments**,^{[7][8][9][10]} (2) **method of maximum likelihood**,^{[7][8][9]} and (3) method of least squares. In the first two methods the interest is in estimating the parameters of the distribution, ν and σ , from a sample of data. This can be done using the method of moments, e.g., the sample mean and the sample standard deviation. The sample mean is an estimate of μ_1 and the sample standard deviation is an estimate of $\mu_2^{1/2}$.

The following is an efficient method, known as the “Koay inversion technique”.^[11] for solving the **estimating equations**, based on the sample mean and the sample standard deviation, simultaneously. This inversion technique is also known as the **fixed point** formula of **SNR**. Earlier works^{[7][12]} on the method of moments usually use a root-finding method to solve the problem, which is not efficient.

First, the ratio of the sample mean to the sample standard deviation is defined as r , i.e., $r = \mu_1' / \mu_2^{1/2}$. The fixed point formula of SNR is expressed as

$$g(\theta) = \sqrt{\xi(\theta) [1 + r^2] - 2},$$

where θ is the ratio of the parameters, i.e., $\theta = \frac{\nu}{\sigma}$, and $\xi(\theta)$ is given by:

$$\xi(\theta) = 2 + \theta^2 - \frac{\pi}{8} \exp(-\theta^2/2) [(2 + \theta^2)I_0(\theta^2/4) + \theta^2 I_1(\theta^2/4)]^2,$$

where I_0 and I_1 are **modified Bessel functions of the first kind**.

Note that $\xi(\theta)$ is a scaling factor of σ and is related to μ_2 by:

$$\mu_2 = \xi(\theta)\sigma^2.$$

To find the fixed point, θ^* , of g , an initial solution is selected, θ_0 , that is greater than the lower bound, which is $\theta_{\text{lowerbound}} = 0$ and occurs when $r = \sqrt{\pi/(4 - \pi)}$ ^[11] (Notice that this is the $r = \mu'_1/\mu_2^{1/2}$ of a Rayleigh distribution). This provides a starting point for the iteration, which uses functional composition, and this continues until $|g^i(\theta_0) - \theta_{i-1}|$ is less than some small positive value. Here, g^i denotes the composition of the same function, g , i times. In practice, we associate the final θ_n for some integer n as the fixed point, θ^* , i.e., $\theta^* = g(\theta^*)$.

Once the fixed point is found, the estimates ν and σ are found through the scaling function, $\xi(\theta)$, as follows:

$$\sigma = \frac{\mu_2^{1/2}}{\sqrt{\xi(\theta^*)}},$$

and

$$\nu = \sqrt{(\mu_1')^2 + (\xi(\theta^*) - 2)\sigma^2}.$$

To speed up the iteration even more, one can use the Newton's method of root-finding. ^[11] This particular approach is highly efficient.

2.32.6 Applications

- The Euclidean norm of a bivariate normally distributed random vector.
- Rician fading
- Effect of sighting error on target shooting. ^[13]

2.32.7 See also

- Rayleigh distribution
- Stephen O. Rice (1907–1986)

2.32.8 Notes

- [1] Liu 2007 (in one of Horn's confluent hypergeometric functions with two variables).
- [2] Annamalai 2000 (in a sum of infinite series).
- [3] Erdelyi 1953.
- [4] Srivastava 1985.
- [5] Richards, M.A., Rice Distribution for RCS, Georgia Institute of Technology (Sep 2006)

- [6] Abramowitz and Stegun (1968) §13.5.1
- [7] Talukdar et al. 1991
- [8] Bonny et al. 1996
- [9] Sijbers et al. 1998
- [10] den Dekker and Sijbers 2014
- [11] Koay et al. 2006 (known as the SNR fixed point formula).
- [12] Abdi 2001
- [13] “Ballistipedia”. Retrieved 4 May 2014.

2.32.9 References

- Abramowitz, M. and Stegun, I. A. (ed.), *Handbook of Mathematical Functions*, National Bureau of Standards, 1964; reprinted Dover Publications, 1965. ISBN 0-486-61272-4
- Rice, S. O., *Mathematical Analysis of Random Noise*. Bell System Technical Journal 24 (1945) 46–156.
- I. Soltani Bozchalooi and Ming Liang (20 November 2007). “A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection”. *Journal of Sound and Vibration* **308** (1–2): 253–254. doi:10.1016/j.jsv.2007.07.038.
- Liu, X. and Hanzo, L., *A Unified Exact BER Performance Analysis of Asynchronous DS-CDMA Systems Using BPSK Modulation over Fading Channels*, IEEE Transactions on Wireless Communications, Volume 6, Issue 10, October 2007, Pages 3504–3509.
- Annamalai, A., Tellambura, C. and Bhargava, V. K., *Equal-Gain Diversity Receiver Performance in Wireless Channels*, IEEE Transactions on Communications, Volume 48, October 2000, Pages 1732–1745.
- Erdelyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F. G., *Higher Transcendental Functions*, Volume 1. McGraw-Hill Book Company Inc., 1953.
- Srivastava, H. M. and Karlsson, P. W., *Multiple Gaussian Hypergeometric Series*. Ellis Horwood Ltd., 1985.
- Sijbers J., den Dekker A. J., Scheunders P. and Van Dyck D., “Maximum Likelihood estimation of Rician distribution parameters”, IEEE Transactions on Medical Imaging, Vol. 17, Nr. 3, p. 357–361, (1998)
- den Dekker, A.J., and Sijbers, J (December 2014). “Data distributions in magnetic resonance images: a review”. *Physica Medica* **30** (7): 725–741. doi:10.1016/j.ejmp.2014.05.002.
- Koay, C.G. and Basser, P. J., *Analytically exact correction scheme for signal extraction from noisy magnitude MR signals*, Journal of Magnetic Resonance, Volume 179, Issue = 2, p. 317–322, (2006)
- Abdi, A., Tepedelenlioglu, C., Kaveh, M., and Giannakis, G. *On the estimation of the K parameter for the Rice fading distribution*, IEEE Communications Letters, Volume 5, Number 3, March 2001, Pages 92–94.
- Talukdar, K.K., and Lawing, William D. (March 1991). “Estimation of the parameters of the Rice distribution”. *Journal of the Acoustical Society of America* **89** (3): 1193–1197. doi:10.1121/1.400532.
- Bonny, J.M., Renou, J.P., and Zanca, M. (November 1996). “Optimal Measurement of Magnitude and Phase from MR Data”. *Journal of Magnetic Resonance, Series B* **113** (2): 136–144. doi:10.1006/jmrb.1996.0166.

2.32.10 External links

- [MATLAB code for Rice/Rician distribution](#) (PDF, mean and variance, and generating random samples)

2.33 Shifted Gompertz distribution

The **shifted Gompertz distribution** is the distribution of the largest of two independent **random variables** one of which has an **exponential distribution** with parameter b and the other has a **Gumbel distribution** with parameters η and b . In its original formulation the distribution was expressed referring to the Gompertz distribution instead of the Gumbel distribution but, since the Gompertz distribution is a reverted Gumbel distribution, the labelling can be considered as accurate. It has been used as a model of **adoption of innovations**. It was proposed by Bemmaor^[1] (1994). Some of its statistical properties have been studied further by Jiménez and Jodrá^[2] (2009).

It has been used to predict the growth and decline of social networks and on-line services and shown to be superior to the Bass model and Weibull distribution (see the work by Christian Bauchhage and co-authors).

2.33.1 Specification

Probability density function

The **probability density function** of the shifted Gompertz distribution is:

$$f(x; b, \eta) = be^{-bx} e^{-\eta e^{-bx}} [1 + \eta (1 - e^{-bx})] \text{ for } x \geq 0.$$

where $b > 0$ is the **scale parameter** and $\eta > 0$ is the **shape parameter** of the shifted Gompertz distribution.

Cumulative distribution function

The **cumulative distribution function** of the shifted Gompertz distribution is:

$$F(x; b, \eta) = (1 - e^{-bx}) e^{-\eta e^{-bx}} \text{ for } x \geq 0.$$

2.33.2 Properties

The shifted Gompertz distribution is right-skewed for all values of η . It is more flexible than the **Gumbel distribution**.

Shapes

The shifted Gompertz density function can take on different shapes depending on the values of the shape parameter η :

- $0 < \eta \leq 0.5$ the probability density function has its mode at 0.
- $\eta > 0.5$ the probability density function has its mode at

$$\text{mode} = -\frac{\ln(z^*)}{b} \quad 0 < z^* < 1$$

where z^* is the smallest root of

$$\eta^2 z^2 - \eta(3 + \eta)z + \eta + 1 = 0,$$

which is

$$z^* = [3 + \eta - (\eta^2 + 2\eta + 5)^{1/2}]/(2\eta).$$

2.33.3 Related distributions

If η varies according to a **gamma distribution** with shape parameter α and scale parameter β (mean = $\alpha\beta$), the distribution of x is Gamma/Shifted Gompertz (G/SG). When α is equal to one, the G/SG reduces to the **Bass model** (Bemmaor 1994). The G/SG has been applied by Dover, Goldenberg and Shapira ^[3](2009) and Van den Bulte and Stremersch ^[4](2004) among others in the context of the diffusion of innovations. The model is discussed in Chandrasekaran and Tellis ^[5](2007).

2.33.4 See also

- Gumbel distribution
- Generalized extreme value distribution
- Mixture model
- Bass model
- Gompertz distribution

2.33.5 References

- [1] Bemmaor, Albert C. (1994). “Modeling the Diffusion of New Durable Goods: Word-of-Mouth Effect Versus Consumer Heterogeneity”. In G. Laurent, G.L. Lilien & B. Pras. *Research Traditions in Marketing*. Boston: Kluwer Academic Publishers. pp. 201–223. ISBN 0-7923-9388-0.
- [2] Jiménez, Fernando; Jodrá, Pedro (2009). “A Note on the Moments and Computer Generation of the Shifted Gompertz Distribution”. *Communications in Statistics - Theory and Methods* **38** (1): 78–89. doi:10.1080/03610920802155502.
- [3] Dover, Yaniv; Goldenberg, Jacob; Shapira, Daniel (2012). “Network Traces on Penetration: Uncovering Degree Distribution From Adoption Data”. *Marketing Science*. doi:10.1287/mksc.1120.0711.
- [4] Van den Bulte, Christophe; Stremersch, Stefan (2004). “Social Contagion and Income Heterogeneity in New Product Diffusion: A Meta-Analytic Test”. *Marketing Science* **23** (4): 530–544. doi:10.1287/mksc.1040.0054.
- [5] Chandrasekaran, Deepa; Tellis, Gerard J. (2007). “A Critical Review of Marketing Research on Diffusion of New Products”. In Naresh K. Malhotra. *Review of Marketing Research* **3**. Armonk: M.E. Sharpe. pp. 39–80. ISBN 978-0-7656-1306-6.

2.34 Type-2 Gumbel distribution

In probability theory, the **Type-2 Gumbel probability density function** is

$$f(x|a, b) = abx^{-a-1}e^{-bx^{-a}}$$

for

$$0 < x < \infty$$

This implies that it is similar to the **Weibull distributions**, substituting $b = \lambda^{-k}$ and $a = -k$. Note however that a positive k (as in the Weibull distribution) would yield a negative a , which is not allowed here as it would yield a negative probability density.

For $0 < a \leq 1$ the mean is infinite. For $0 < a \leq 2$ the variance is infinite.

The cumulative distribution function is

$$F(x|a, b) = e^{-bx^{-a}}$$

The moments $E[X^k]$ exist for $k < a$

The special case $b = 1$ yields the **Fréchet distribution**

Based on **The GNU Scientific Library**, used under GFDL.

2.34.1 See also

- Extreme value theory
- Gumbel distribution
- Type-1 Gumbel distribution

2.35 Weibull distribution

In probability theory and statistics, the **Weibull distribution** /ˈweɪbəl/ is a continuous probability distribution. It is named after Swedish mathematician **Waloddi Weibull**, who described it in detail in 1951, although it was first identified by **Fréchet** (1927) and first applied by **Rosin & Rammler** (1933) to describe a **particle size distribution**.

2.35.1 Definition

The probability density function of a Weibull random variable is:^[1]

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

where $k > 0$ is the *shape parameter* and $\lambda > 0$ is the *scale parameter* of the distribution. Its complementary cumulative distribution function is a **stretched exponential function**. The Weibull distribution is related to a number of other probability distributions; in particular, it interpolates between the **exponential distribution** ($k = 1$) and the **Rayleigh distribution** ($k = 2$ and $\lambda = \sqrt{2}\sigma$ ^[2]).

If the quantity X is a “time-to-failure”, the Weibull distribution gives a distribution for which the **failure rate** is proportional to a power of time. The *shape* parameter, k , is that power plus one, and so this parameter can be interpreted directly as follows:

- A value of $k < 1$ indicates that the failure rate decreases over time. This happens if there is significant “infant mortality”, or defective items failing early and the failure rate decreasing over time as the defective items are weeded out of the population.
- A value of $k = 1$ indicates that the failure rate is constant over time. This might suggest random external events are causing mortality, or failure.
- A value of $k > 1$ indicates that the failure rate increases with time. This happens if there is an “aging” process, or parts that are more likely to fail as time goes on.

In the field of materials science, the shape parameter k of a distribution of strengths is known as the **Weibull modulus**.

2.35.2 Properties

Density function

The form of the density function of the Weibull distribution changes drastically with the value of k . For $0 < k < 1$, the density function tends to ∞ as x approaches zero from above and is strictly decreasing. For $k = 1$, the density function tends to $1/\lambda$ as x approaches zero from above and is strictly decreasing. For $k > 1$, the density function tends to zero as x approaches zero from above, increases until its mode and decreases after it. It is interesting to note that the density function has infinite negative slope at $x = 0$ if $0 < k < 1$, infinite positive slope at $x = 0$ if $1 < k < 2$ and null slope at $x = 0$ if $k > 2$. For $k = 2$ the density has a finite positive slope at $x = 0$. As k goes to infinity, the Weibull distribution converges to a **Dirac delta distribution** centered at $x = \lambda$. Moreover, the skewness and coefficient of variation depend only on the shape parameter.

Distribution function

The **cumulative distribution function** for the Weibull distribution is

$$F(x; k, \lambda) = 1 - e^{-(x/\lambda)^k}$$

for $x \geq 0$, and $F(x; k, \lambda) = 0$ for $x < 0$.

The quantile (inverse cumulative distribution) function for the Weibull distribution is

$$Q(p; k, \lambda) = \lambda(-\ln(1 - p))^{1/k}$$

for $0 \leq p < 1$.

The **failure rate** h (or hazard function) is given by

$$h(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1}.$$

Moments

The **moment generating function** of the **logarithm** of a Weibull distributed **random variable** is given by^[3]

$$E[e^{t \log X}] = \lambda^t \Gamma\left(\frac{t}{k} + 1\right)$$

where Γ is the **gamma function**. Similarly, the **characteristic function** of $\log X$ is given by

$$E[e^{it \log X}] = \lambda^{it} \Gamma\left(\frac{it}{k} + 1\right).$$

In particular, the n th **raw moment** of X is given by

$$m_n = \lambda^n \Gamma\left(1 + \frac{n}{k}\right).$$

The mean and variance of a Weibull random variable can be expressed as

$$E(X) = \lambda \Gamma \left(1 + \frac{1}{k} \right)$$

and

$$\text{var}(X) = \lambda^2 \left[\Gamma \left(1 + \frac{2}{k} \right) - \left(\Gamma \left(1 + \frac{1}{k} \right) \right)^2 \right].$$

The skewness is given by

$$\gamma_1 = \frac{\Gamma \left(1 + \frac{3}{k} \right) \lambda^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

where the mean is denoted by μ and the standard deviation is denoted by σ .

The excess **kurtosis** is given by

$$\gamma_2 = \frac{-6\Gamma_1^4 + 12\Gamma_1^2\Gamma_2 - 3\Gamma_2^2 - 4\Gamma_1\Gamma_3 + \Gamma_4}{[\Gamma_2 - \Gamma_1^2]^2}$$

where $\Gamma_i = \Gamma(1 + i/k)$. The kurtosis excess may also be written as:

$$\gamma_2 = \frac{\lambda^4 \Gamma(1 + \frac{4}{k}) - 4\gamma_1 \sigma^3 \mu - 6\mu^2 \sigma^2 - \mu^4}{\sigma^4} - 3$$

Moment generating function

A variety of expressions are available for the moment generating function of X itself. As a **power series**, since the raw moments are already known, one has

$$E[e^{tX}] = \sum_{n=0}^{\infty} \frac{t^n \lambda^n}{n!} \Gamma \left(1 + \frac{n}{k} \right).$$

Alternatively, one can attempt to deal directly with the integral

$$E[e^{tX}] = \int_0^{\infty} e^{tx} \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} e^{-(x/\lambda)^k} dx.$$

If the parameter k is assumed to be a rational number, expressed as $k = p/q$ where p and q are integers, then this integral can be evaluated analytically.^[4] With t replaced by $-t$, one finds

$$E[e^{-tX}] = \frac{1}{\lambda^k t^k} \frac{p^k \sqrt{q/p}}{(\sqrt{2\pi})^{q+p-2}} G_{p,q}^{q,p} \left(\frac{1-k}{p}, \frac{2-k}{p}, \dots, \frac{p-k}{p} \mid \frac{p^p}{(q \lambda^k t^k)^q} \right)$$

where G is the **Meijer G-function**.

The characteristic function has also been obtained by **Muraleedharan et al. (2007)**. The characteristic function and moment generating function of 3-parameter Weibull distribution have also been derived by **Muraleedharan & Soares (2014)** by a direct approach.

Information entropy

The information entropy is given by

$$H(\lambda, k) = \gamma \left(1 - \frac{1}{k}\right) + \ln \left(\frac{\lambda}{k}\right) + 1$$

where γ is the Euler–Mascheroni constant.

Parameter estimation

Maximum likelihood The maximum likelihood estimator for the λ parameter given k is,

$$\hat{\lambda}^k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

The maximum likelihood estimator for k is,

$$\hat{k}^{-1} = \frac{\sum_{i=1}^n x_i^k \ln x_i}{\sum_{i=1}^n x_i^k} - \frac{1}{n} \sum_{i=1}^n \ln x_i$$

This being an implicit function, one must generally solve for k by numerical means.

When $x_1 > x_2 > \dots > x_N$ are the N largest observed samples from a dataset of more than N samples, then the maximum likelihood estimator for the λ parameter given k is,^[5]

$$\hat{\lambda}^k = \frac{1}{N} \sum_{i=1}^N (x_i^k - x_N^k)$$

Also given that condition, the maximum likelihood estimator for k is,

$$\hat{k}^{-1} = \frac{\sum_{i=1}^N (x_i^k \ln x_i - x_N^k \ln x_N)}{\sum_{i=1}^N (x_i^k - x_N^k)} - \frac{1}{N} \sum_{i=1}^N \ln x_i$$

Again, this being an implicit function, one must generally solve for k by numerical means.

2.35.3 Weibull plot

The fit of data to a Weibull distribution can be visually assessed using a Weibull Plot.^[6] The Weibull Plot is a plot of the empirical cumulative distribution function $\hat{F}(x)$ of data on special axes in a type of Q-Q plot. The axes are $\ln(-\ln(1 - \hat{F}(x)))$ versus $\ln(x)$. The reason for this change of variables is the cumulative distribution function can be linearized:

$$\begin{aligned} F(x) &= 1 - e^{-(x/\lambda)^k} \\ -\ln(1 - F(x)) &= (x/\lambda)^k \\ \underbrace{\ln(-\ln(1 - F(x)))}_{\text{'y'}} &= \underbrace{k \ln x}_{\text{'mx'}} - \underbrace{k \ln \lambda}_{\text{'c'}} \end{aligned}$$

which can be seen to be in the standard form of a straight line. Therefore if the data came from a Weibull distribution then a straight line is expected on a Weibull plot.

There are various approaches to obtaining the empirical distribution function from data: one method is to obtain the vertical coordinate for each point using $\hat{F} = \frac{i-0.3}{n+0.4}$ where i is the rank of the data point and n is the number of data points.^[7]

Linear regression can also be used to numerically assess goodness of fit and estimate the parameters of the Weibull distribution. The gradient informs one directly about the shape parameter k and the scale parameter λ can also be inferred.

The Weibull distribution is used

- In survival analysis^[8]
- In reliability engineering and failure analysis
- In electrical engineering to represent overvoltage occurring in an electrical system
- In industrial engineering to represent manufacturing and delivery times
- In extreme value theory
- In weather forecasting
 - To describe wind speed distributions, as the natural distribution often matches the Weibull shape^[9]
- In communications systems engineering
 - In radar systems to model the dispersion of the received signals level produced by some types of clutters
 - To model fading channels in wireless communications, as the Weibull fading model seems to exhibit good fit to experimental fading channel measurements
- In general insurance to model the size of reinsurance claims, and the cumulative development of asbestosis losses
- In forecasting technological change (also known as the Sharif-Islam model)^[10]
- In hydrology the Weibull distribution is applied to extreme events such as annual maximum one-day rainfalls and river discharges. The blue picture illustrates an example of fitting the Weibull distribution to ranked annually maximum one-day rainfalls showing also the 90% confidence belt based on the binomial distribution. The rainfall data are represented by plotting positions as part of the cumulative frequency analysis.
- In describing the size of particles generated by grinding, milling and crushing operations, the 2-Parameter Weibull distribution is used, and in these applications it is sometimes known as the Rosin-Rammler distribution. In this context it predicts fewer fine particles than the Log-normal distribution and it is generally most accurate for narrow particle size distributions. The interpretation of the cumulative distribution function is that $F(x; k; \lambda)$ is the mass fraction of particles with diameter smaller than x , where λ is the mean particle size and k is a measure of the spread of particle sizes.

2.35.4 Related distributions

- The translated Weibull distribution (or 3-parameter Weibull) contains an additional parameter.^[3] It has the probability density function

$$f(x; k, \lambda, \theta) = \frac{k}{\lambda} \left(\frac{x - \theta}{\lambda} \right)^{k-1} e^{-\left(\frac{x-\theta}{\lambda}\right)^k}$$

for $x \geq \theta$ and $f(x; k, \lambda, \theta) = 0$ for $x < \theta$, where $k > 0$ is the shape parameter, $\lambda > 0$ is the scale parameter and θ is the location parameter of the distribution. When $\theta=0$, this reduces to the 2-parameter distribution.

- The Weibull distribution can be characterized as the distribution of a random variable W such that the random variable

$$X = \left(\frac{W}{\lambda} \right)^k$$

is the standard **exponential distribution** with intensity 1.^[3]

- This implies that the Weibull distribution can also be characterized in terms of a **uniform distribution**: if U is uniformly distributed on $(0,1)$, then the random variable $W = \lambda(-\ln(U))^{1/k}$ is Weibull distributed with parameters k and λ . (Note that $-\ln(U)$ here is equivalent to X just above.) This leads to an easily implemented numerical scheme for simulating a Weibull distribution.
- The Weibull distribution interpolates between the exponential distribution with intensity $1/\lambda$ when $k = 1$ and a **Rayleigh distribution** of mode $\sigma = \lambda/\sqrt{2}$ when $k = 2$.
- The Weibull distribution (usually sufficient in **reliability engineering**) is a special case of the three parameter **exponentiated Weibull distribution** where the additional exponent equals 1. The exponentiated Weibull distribution accommodates **unimodal, bathtub shaped**^[11] and **monotone failure rates**.
- The Weibull distribution is a special case of the **generalized extreme value distribution**. It was in this connection that the distribution was first identified by **Maurice Fréchet** in 1927.^[12] The closely related **Fréchet distribution**, named for this work, has the probability density function

$$f_{\text{Fréchet}}(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{-1-k} e^{-(x/\lambda)^{-k}} = *f_{\text{Weibull}}(x; -k, \lambda).$$

- The distribution of a random variable that is defined as the minimum of several random variables, each having a different Weibull distribution, is a **poly-Weibull distribution**.
- The Weibull distribution was first applied by **Rosin & Rammler (1933)** to describe particle size distributions. It is widely used in **mineral processing** to describe **particle size distributions** in **comminution processes**. In this context the cumulative distribution is given by

$$f(x; P_{80}, m) = \begin{cases} 1 - e^{\ln(0.2) \left(\frac{x}{P_{80}} \right)^m} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

where

x

P_{80}

m

- Because of its availability in **Spreadsheets**, it is also used where the underlying behavior is actually better modeled by an **Erlang distribution**.^[13]

2.35.5 See also

- **Fisher–Tippett–Gnedenko theorem**
- **Logistic distribution**
- **Rosin–Rammler distribution** for particle size analysis

2.35.6 References

- [1] Papoulis, Athanasios Papoulis; Pillai, S. Unnikrishna (2002). *Probability, Random Variables, and Stochastic Processes* (4th ed.). Boston: McGraw-Hill. ISBN 0-07-366011-6.
- [2] <http://www.mathworks.com.au/help/stats/rayleigh-distribution.html>
- [3] Johnson, Kotz & Balakrishnan 1994
- [4] See (Cheng, Tellambura & Beaulieu 2004) for the case when k is an integer, and (Sagias & Karagiannidis 2005) for the rational case.
- [5] Sornette, D. (2004). *Critical Phenomena in Natural Science: Chaos, Fractals, Self-organization, and Disorder..*
- [6] The Weibull plot
- [7] Wayne Nelson (2004) *Applied Life Data Analysis*. Wiley-Blackwell ISBN 0-471-64462-5
- [8] Survival/Failure Time Analysis
- [9] Wind Speed Distribution Weibull
- [10] “The Weibull distribution as a general model for forecasting technological change”. *Technological Forecasting and Social Change* **18**: 247–256. doi:10.1016/0040-1625(80)90026-8. Retrieved 2013-09-05.
- [11] “System evolution and reliability of systems”. Sysev (Belgium). 2010-01-01.
- [12] Montgomery, Douglas. *Introduction to statistical quality control*. [S.l.]: John Wiley. p. 95. ISBN 9781118146811.
- [13] Chatfield, C.; Goodhardt, G.J. (1973). “A Consumer Purchasing Model with Erlang Interpurchase Times”. *Journal of the American Statistical Association* **68**: 828–835. doi:10.1080/01621459.1973.10481432.

2.35.7 Bibliography

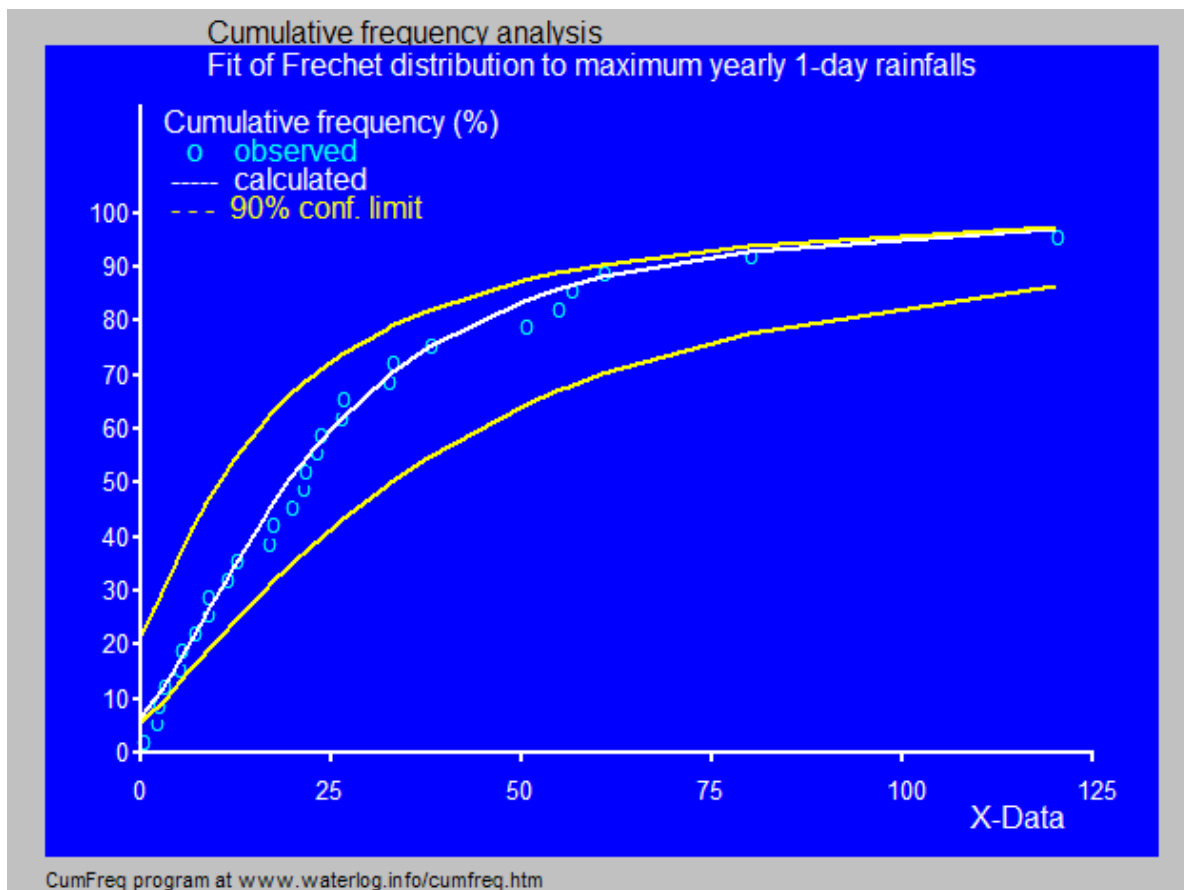
- Fréchet, Maurice (1927), “Sur la loi de probabilité de l'écart maximum”, *Annales de la Société Polonaise de Mathématique, Cracovie* **6**: 93–116.
- Johnson, Norman L.; Kotz, Samuel; Balakrishnan, N. (1994), *Continuous univariate distributions. Vol. 1*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (2nd ed.), New York: John Wiley & Sons, ISBN 978-0-471-58495-7, MR 1299979
- Muraleedharan, G.; Rao, A.D.; Kurup, P.G.; Nair, N. Unnikrishnan; Sinha, Mourani (2007), “Modified Weibull Distribution for Maximum and Significant Wave Height Simulation and Prediction”, *Coastal Engineering* **54** (8): 630–638, doi:10.1016/j.coastaleng.2007.05.001
- Muraleedharan, G.; Soares, C.G. (2014), “Characteristic and Moment Generating Functions of Generalised Pareto (GP3) and Weibull Distributions”, *Journal of Scientific Research and Reports* **3** (14): 1861–1874, doi:10.9734/JSRR/2014/10087.
- Rosin, P.; Rammler, E. (1933), “The Laws Governing the Fineness of Powdered Coal”, *Journal of the Institute of Fuel* **7**: 29–36.
- Sagias, Nikos C.; Karagiannidis, George K. (2005), “Gaussian class multivariate Weibull distributions: theory and applications in fading channels” (PDF), *Institute of Electrical and Electronics Engineers. Transactions on Information Theory* **51** (10): 3608–3619, doi:10.1109/TIT.2005.855598, ISSN 0018-9448, MR 2237527
- Weibull, W. (1951), “A statistical distribution function of wide applicability” (PDF), *J. Appl. Mech.-Trans. ASME* **18** (3): 293–297.
- “Engineering statistics handbook”. National Institute of Standards and Technology. 2008. |chapter= ignored (help)
- Nelson, Jr, Ralph (2008-02-05). “Dispersing Powders in Liquids, Part 1, Chap 6: Particle Volume Distribution”. Retrieved 2008-02-05.

2.35.8 External links

- Hazewinkel, Michiel, ed. (2001), “Weibull distribution”, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Mathpages - Weibull Analysis
- The Weibull Distribution
- Reliability Analysis with Weibull
- Interactive graphic: Univariate Distribution Relationships



Ronald Fisher



Fitted cumulative Fréchet distribution to extreme one-day rainfalls

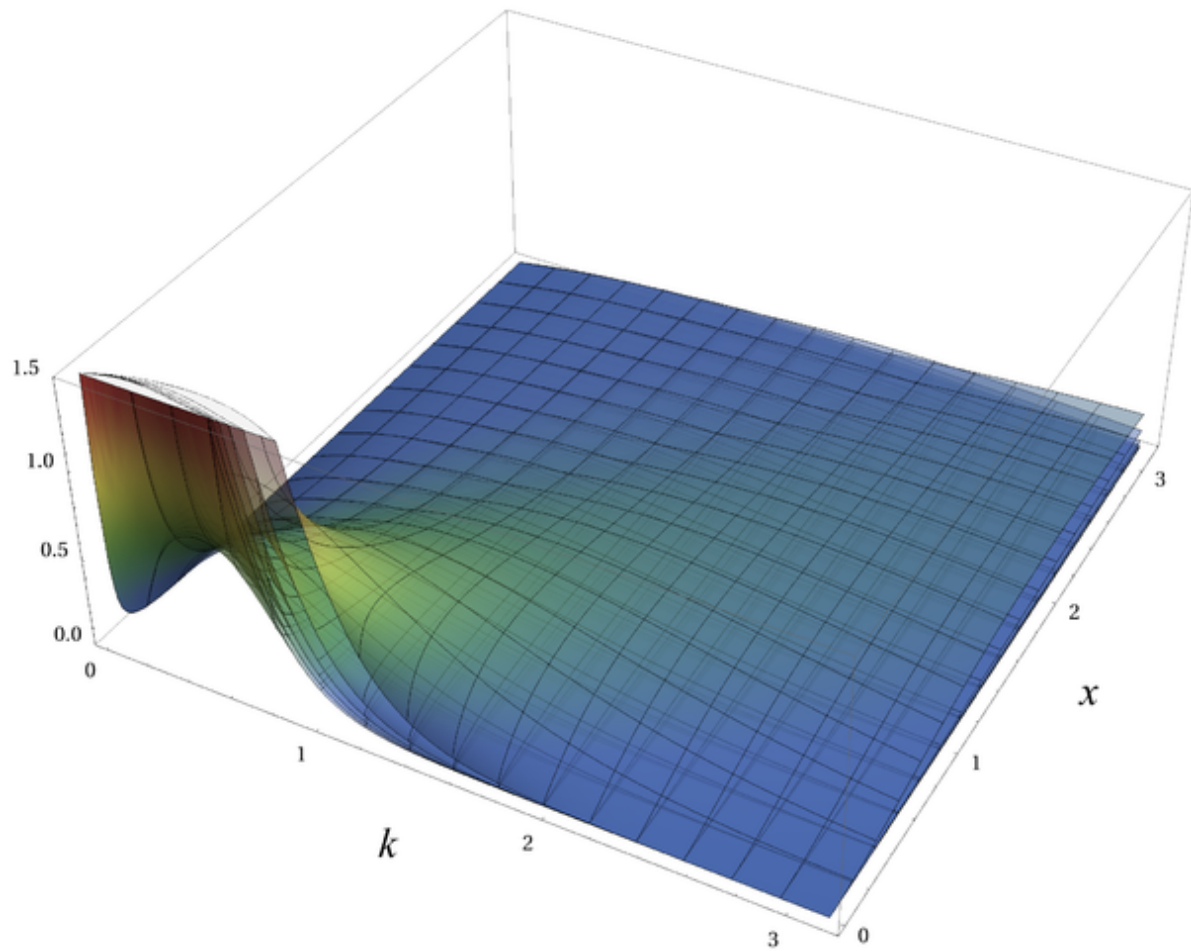


Illustration of the gamma PDF for parameter values over k and x with θ set to 1, 2, 3, 4, 5 and 6. One can see each θ layer by itself here as well as by k and x .

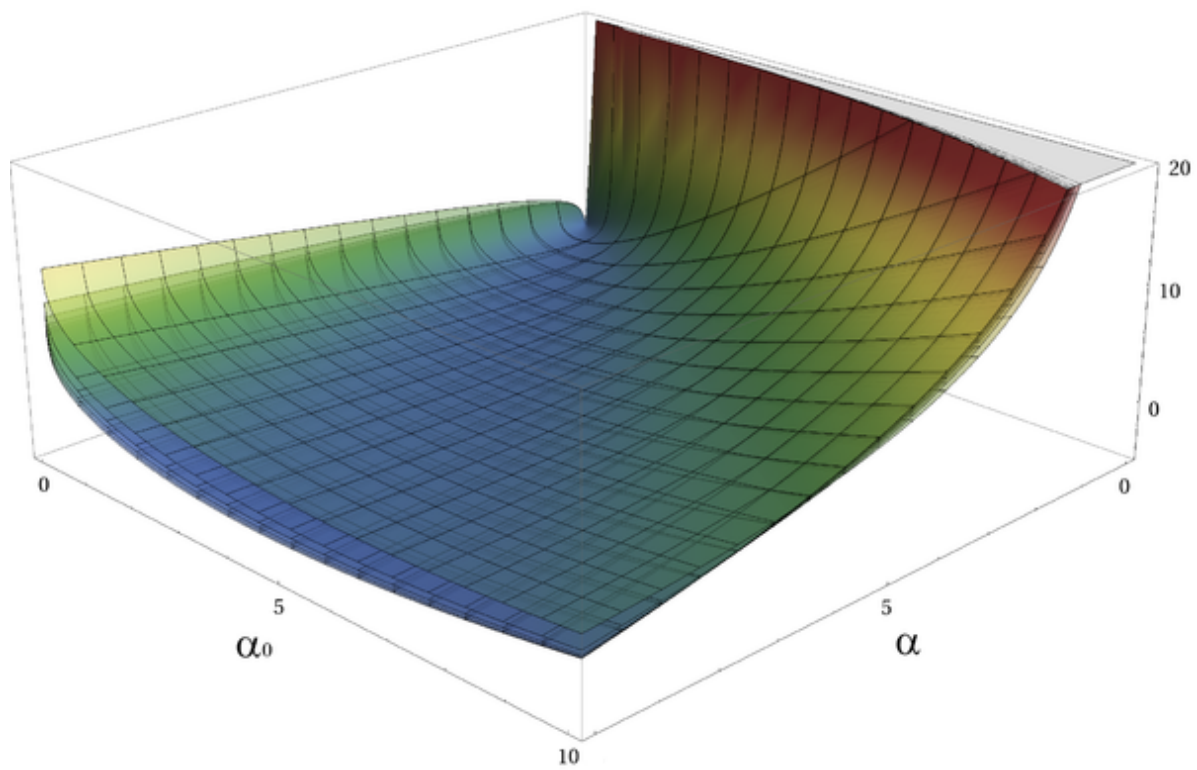
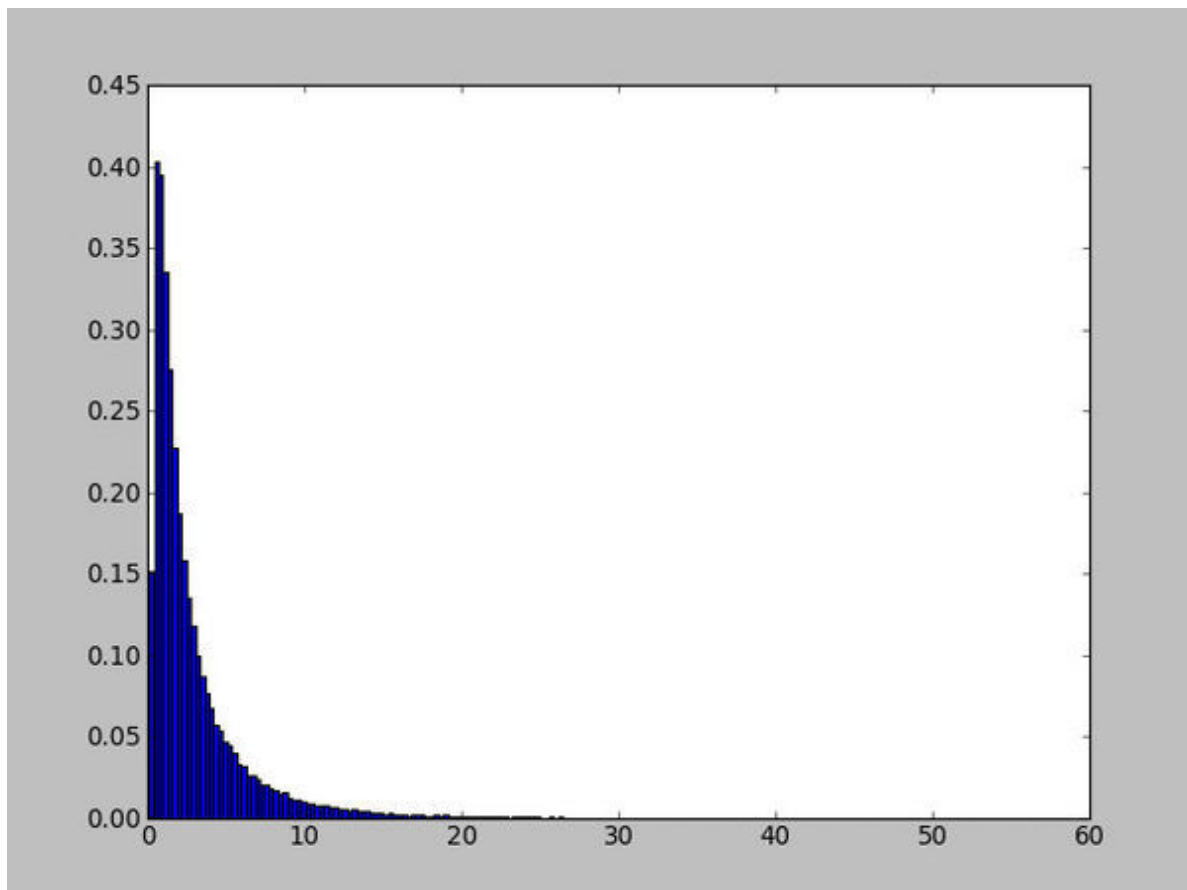
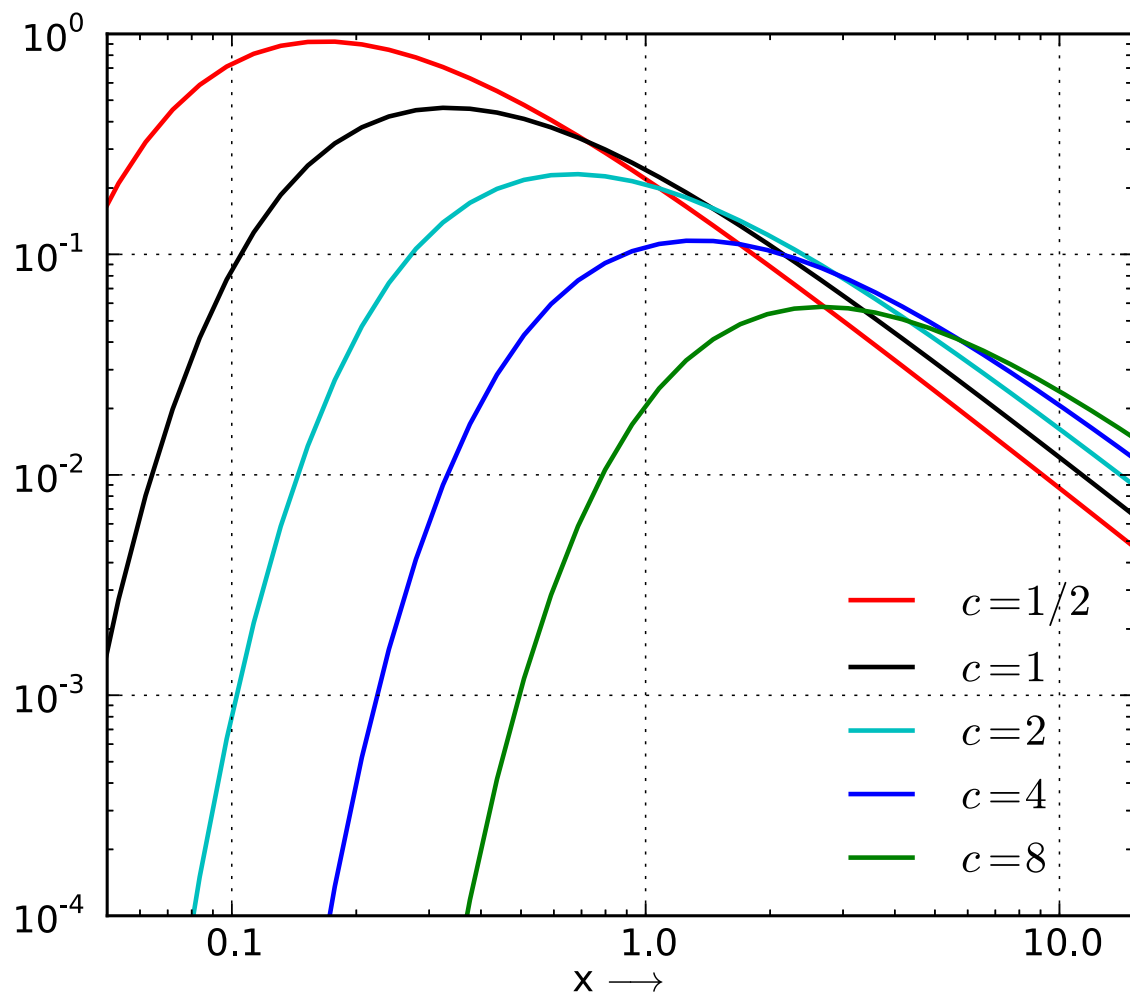


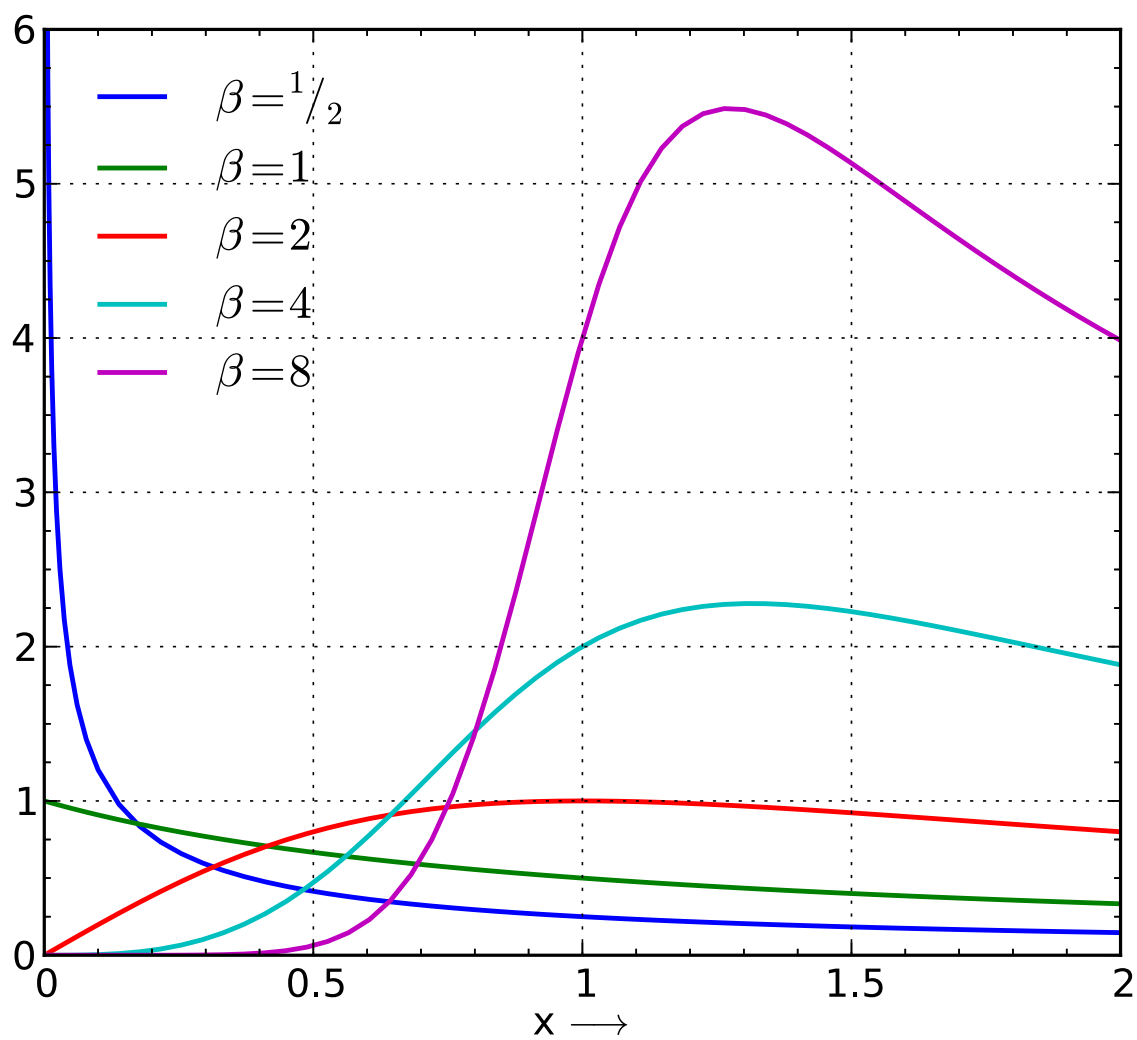
Illustration of the Kullback–Leibler (KL) divergence for two gamma PDFs. Here $\beta = \beta_0 + 1$ which are set to 1, 2, 3, 4, 5 and 6. The typical asymmetry for the KL divergence is clearly visible.



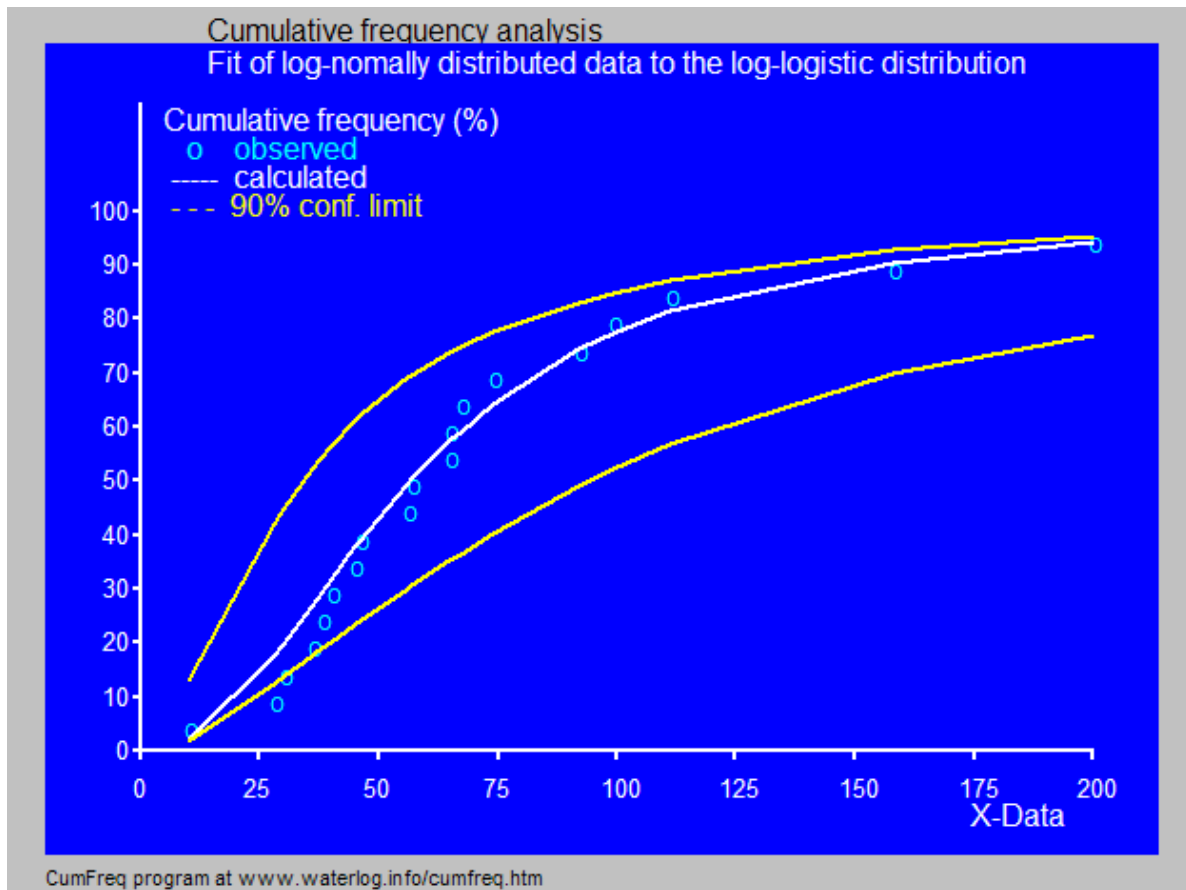
Wald Distribution using Python with aid of matplotlib and NumPy



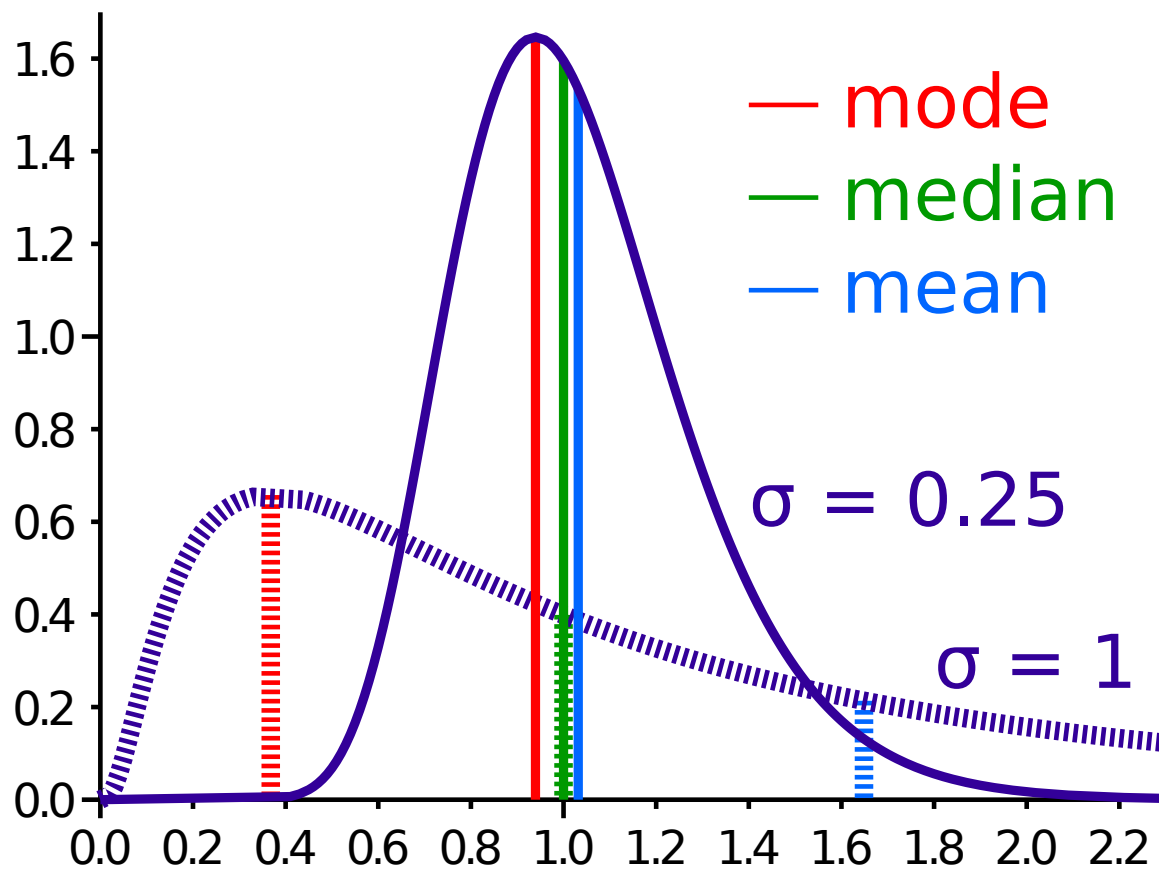
Probability density function for the Lévy distribution on a log-log scale.



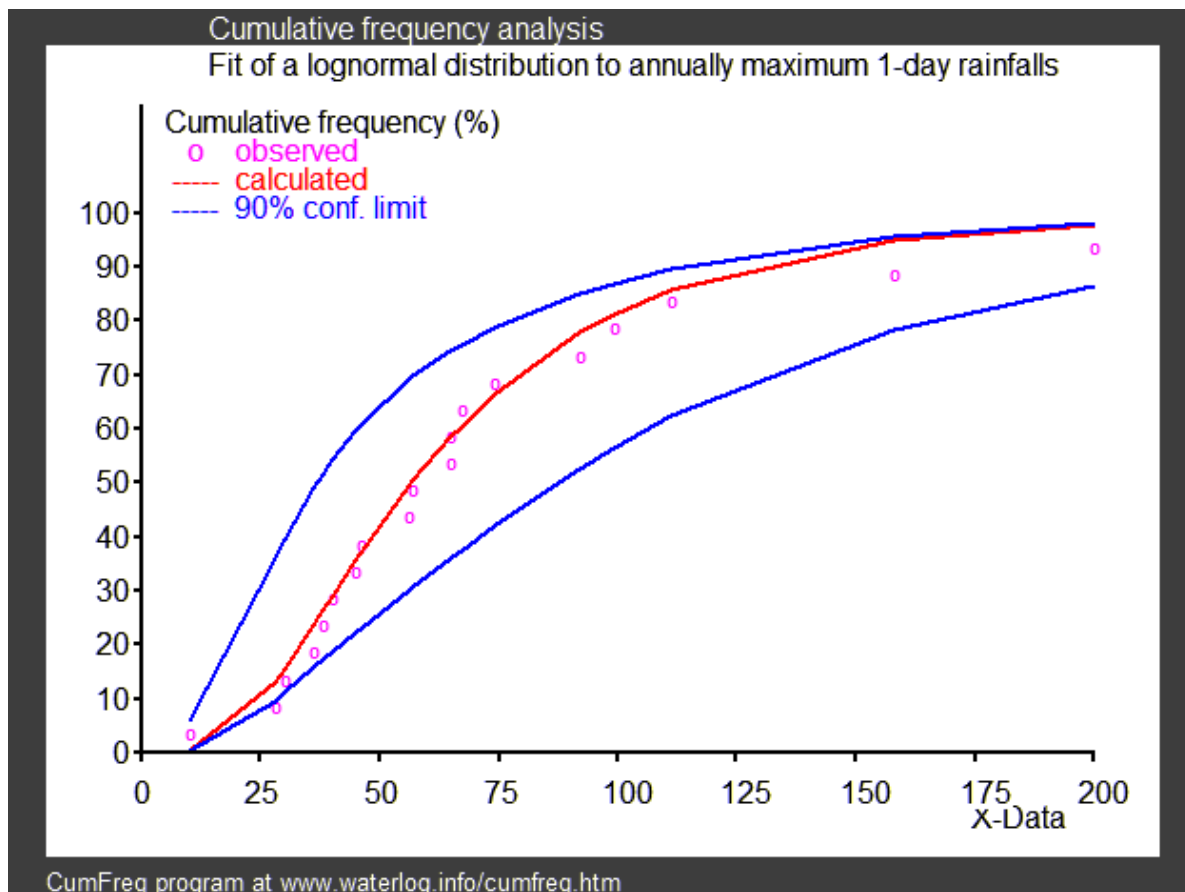
Hazard function. $\alpha = 1$, values of β as shown in legend



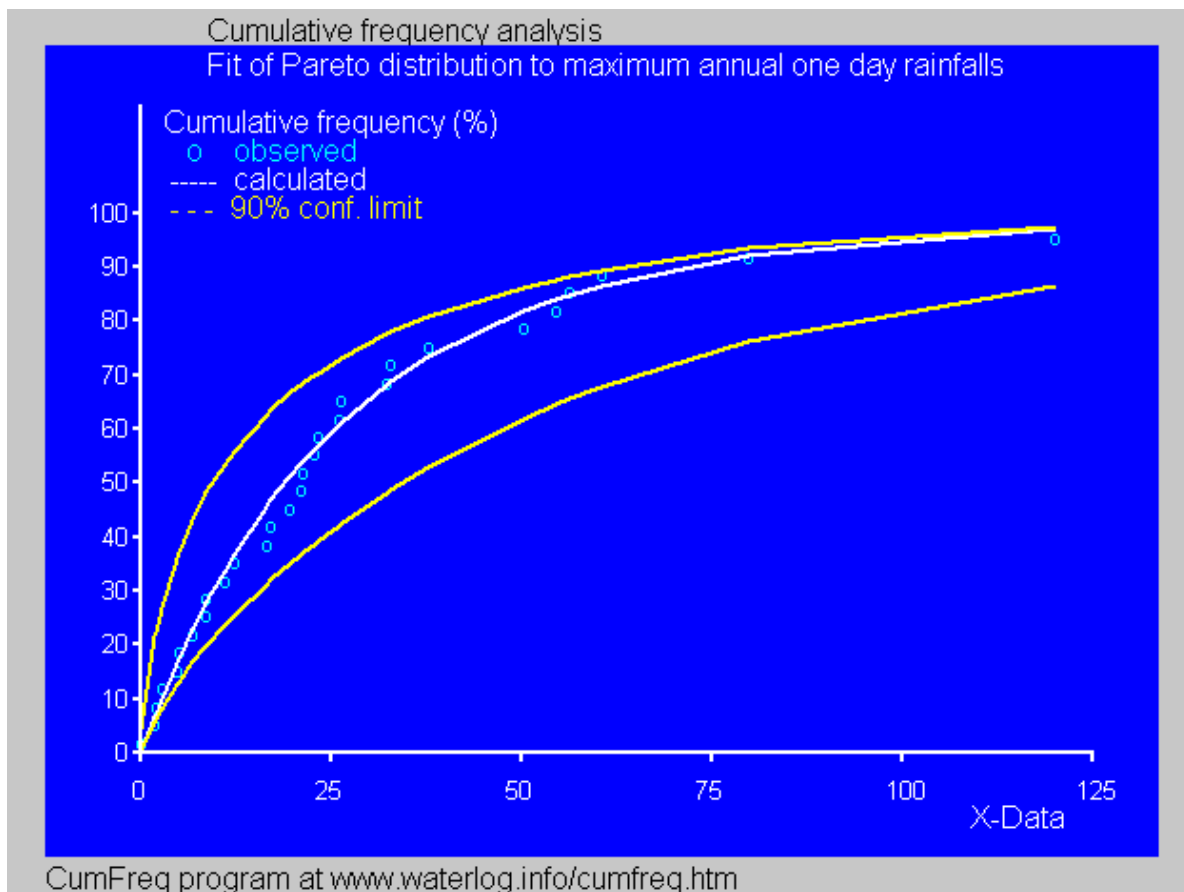
Fitted cumulative log-logistic distribution to maximum one-day October rainfalls using CumFreq, see also Distribution fitting



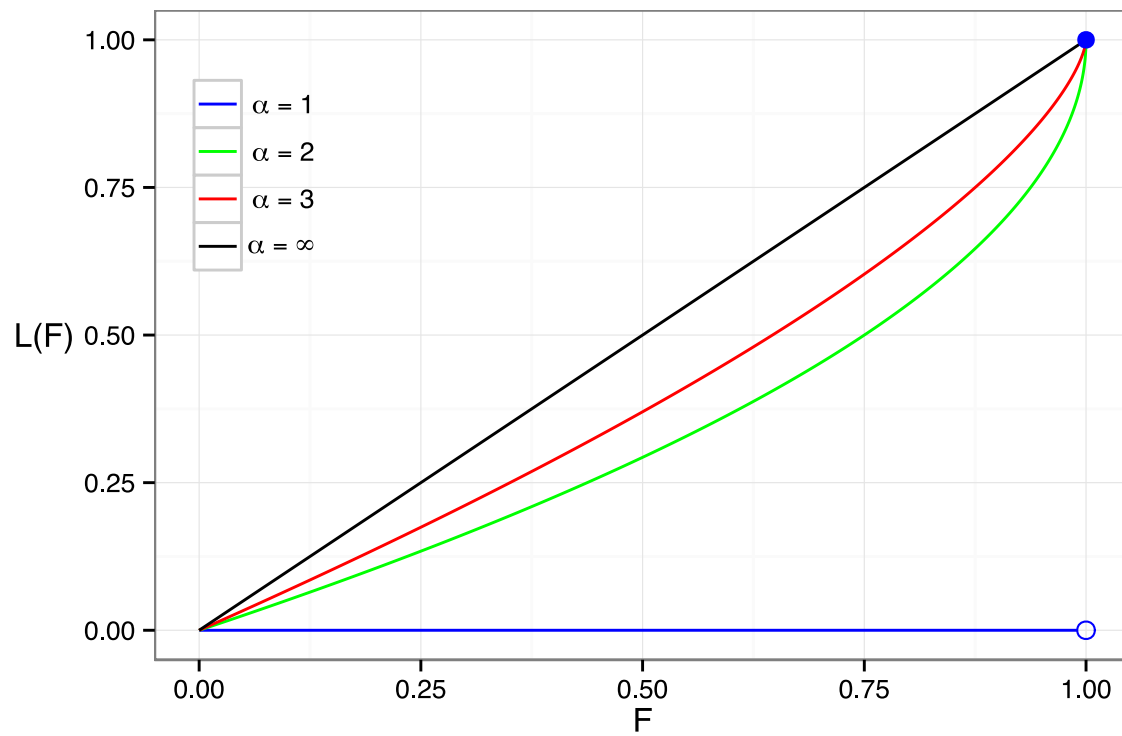
Comparison of mean, median and mode of two log-normal distributions with different skewness.



Fitted cumulative log-normal distribution to annually maximum 1-day rainfalls, see distribution fitting



Fitted cumulative Pareto (Lomax) distribution to maximum one-day rainfalls using CumFreq, see also distribution fitting



Lorenz curves for a number of Pareto distributions. The case $\alpha = \infty$ corresponds to perfectly equal distribution ($G = 0$) and the $\alpha = 1$ line corresponds to complete inequality ($G = 1$)

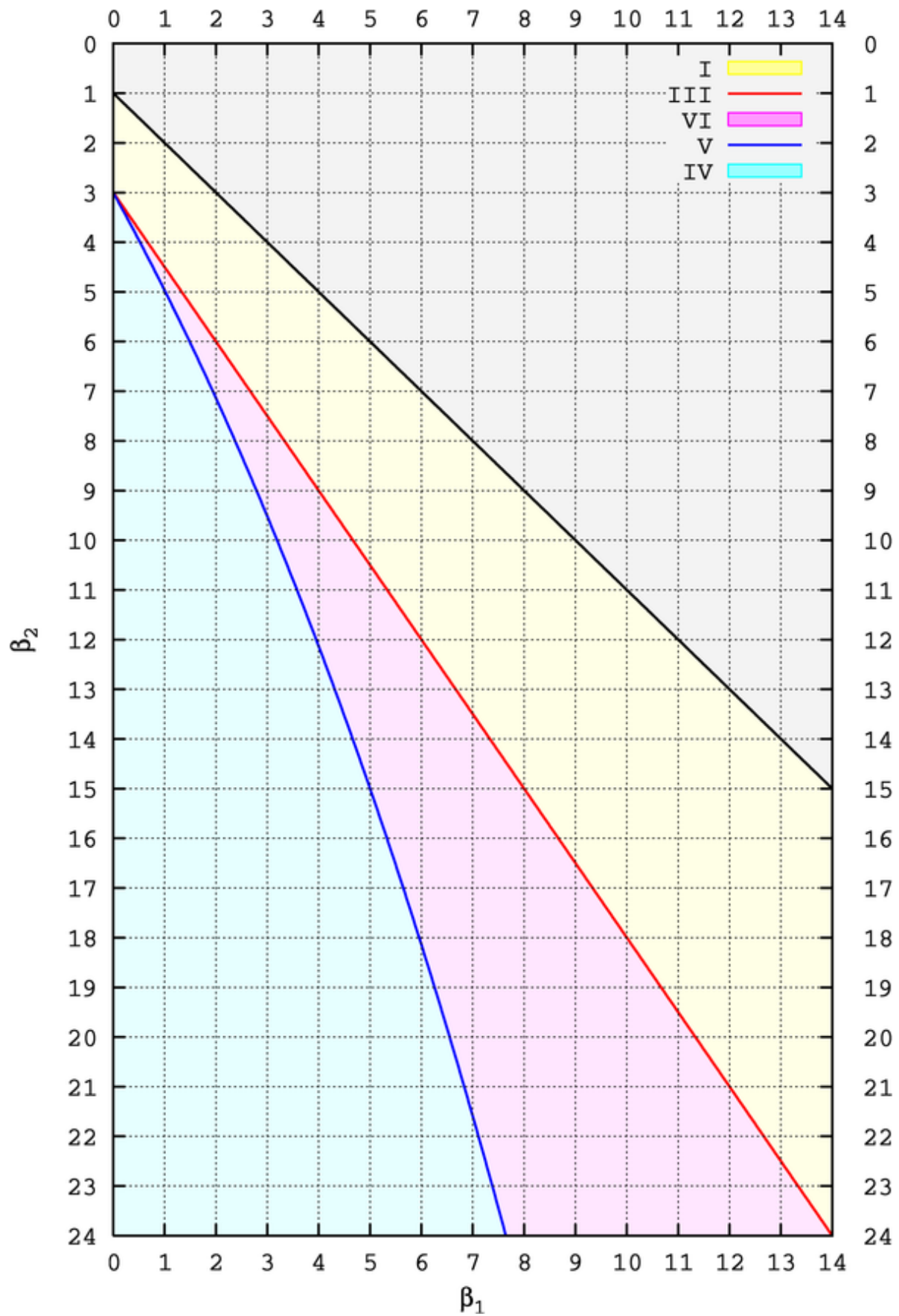
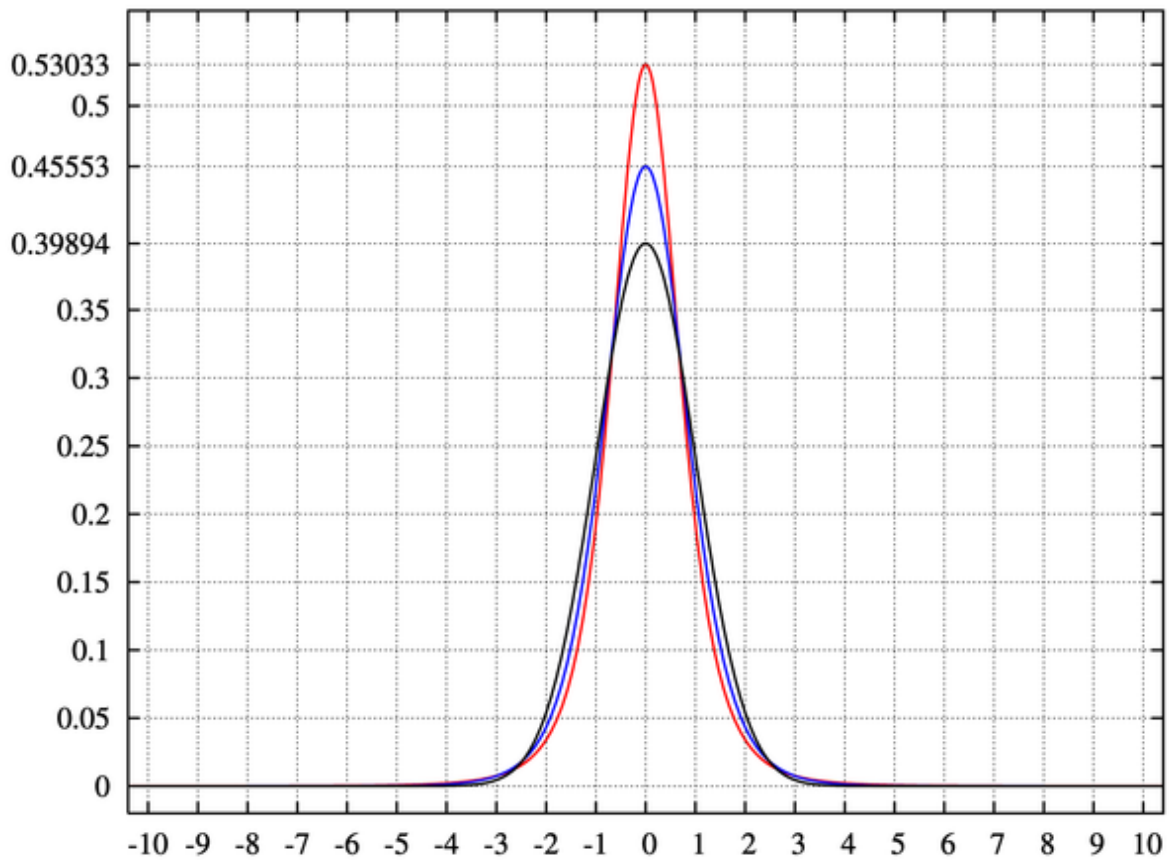
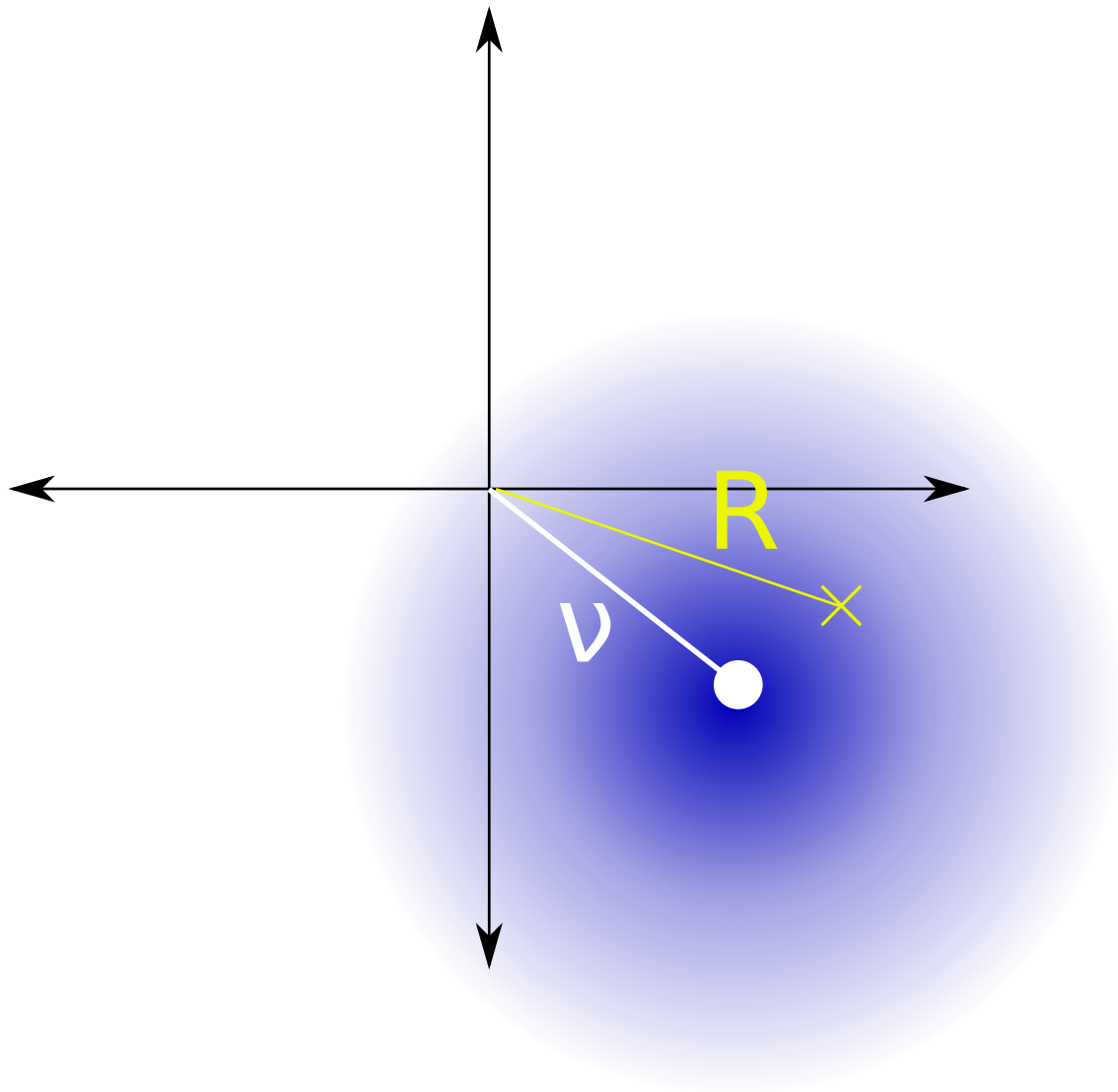


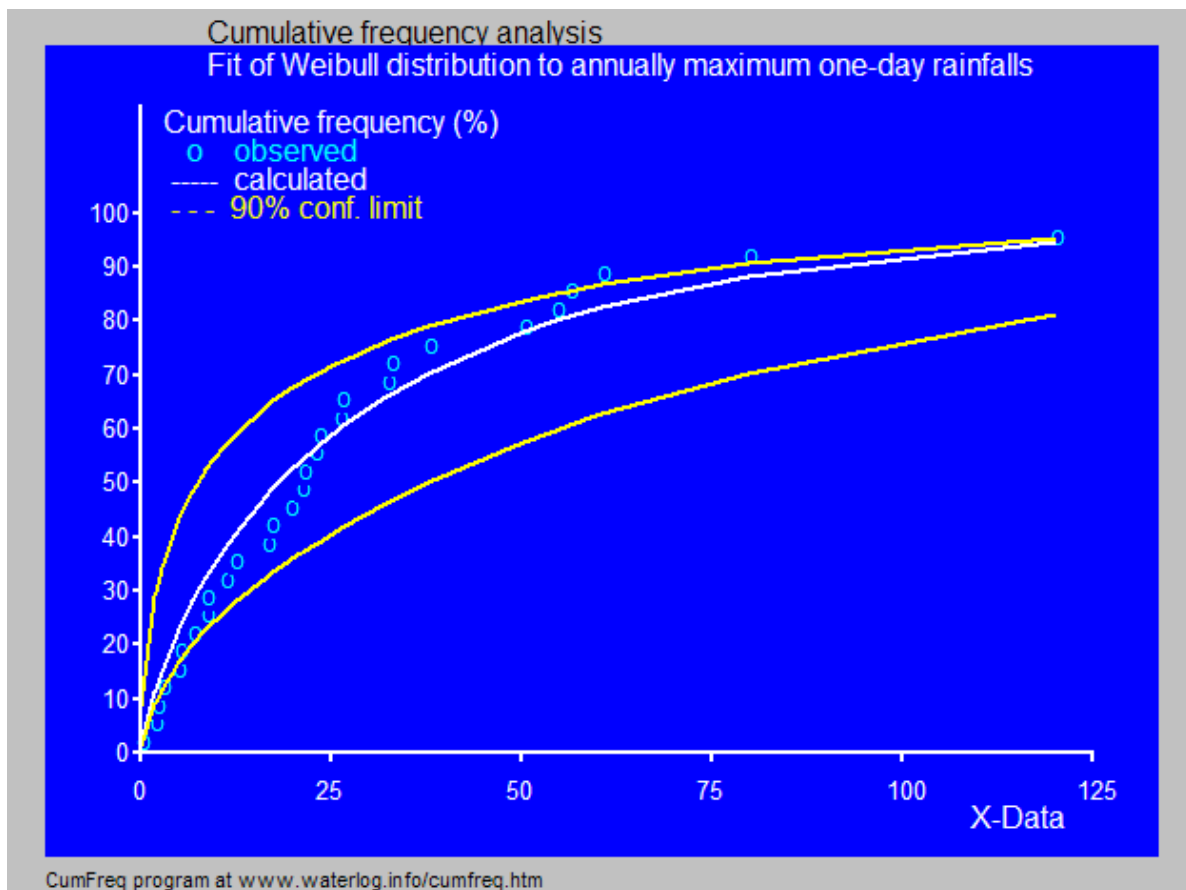
Diagram of the Pearson system, showing distributions of types I, III, VI, V, and IV in terms of β_1 (squared skewness) and β_2 (traditional kurtosis)



Plot of Pearson type VII densities with $\lambda = 0$, $\sigma = 1$, and: $\gamma_2 = \infty$ (red); $\gamma_2 = 4$ (blue); and $\gamma_2 = 0$ (black)



In the 2D plane, pick a fixed point at distance v from the origin. Generate a distribution of 2D points centered around that point, where the x and y coordinates are chosen independently from a gaussian distribution with standard deviation σ (blue region). If R is the distance from these points to the origin, then R has a Rice distribution.



Fitted cumulative Weibull distribution to maximum one-day rainfalls using CumFreq, see also distribution fitting

Chapter 3

Text and image sources, contributors, and licenses

3.1 Text

- **List of probability distributions** *Source:* https://en.wikipedia.org/wiki/List_of_probability_distributions?oldid=683689286 *Contributors:* Fnielsen, Michael Hardy, Robinh, Giftlite, Smalljim, PAR, Mindmatrix, Btyner, Nanite, Shawn@garbett.org, Schmock, Gargoyle888, Katieh5584, Fuzzyrandom, Oli Filth, Kjetil1001, G716, Autopilot, Ben Moore, Peleg, Winterfors, Cydebot, Davidhof, IanOsgood, Coffee2theorems, Albmont, Paresnah, It Is Me Here, Wastle, DrMicro, TXiKiBoT, Nschuma, Rlendog, Sheppa28, Yerpo, Melcombe, Rumping, Carolus m, Skbkekass, Qwfp, Addbot, Rmalouf, Cbauckhage, LinkFA-Bot, Mdnahas, Csigabi, Xqbot, Tomaschwutz, Ehsan.azhdari, Kastchei, Pjsanchez, Ibayes, GonzoEcon, Illia Connell, Dexbot, MoreHumanThanNot, Smason79, Limit-theorem, Parwig, Herbmuell, Julia Abril, Carterkd and Anonymous: 30
- **Bernoulli distribution** *Source:* https://en.wikipedia.org/wiki/Bernoulli_distribution?oldid=689855034 *Contributors:* Bryan Derksen, Miguel~enwiki, Olivier, Michael Hardy, Tomi, TakuyaMurata, Whkoh, Poor Yorick, Rossami, Charles Matthews, Jitse Niesen, Robbot, Wtanaka, Weialawaga~enwiki, Giftlite, MarkSweep, Rdsmith4, Urhixidur, Discospinster, Jpk, El C, Runner1928, Musiphil, Eric Kvaalen, Complex01, PAR, Cburnett, Aquae, Tomash, Btyner, Mathbot, YurikBot, Wavelength, Schmock, SmackBot, RDBury, Federalist51, Zven, Iwaterpolo, Bando26, Yoderj, FilipeS, AlekseyP, Thijs!bot, Wikid77, Lovibond, Pabristow, .anacondabot, Albmont, Aziz1005, User A1, Lilac Soul, ILikeHowMuch, Policron, Philip Trueman, TXiKiBoT, A4bot, Camkego, Typofier, Sharmistha1, OKBot, Melcombe, Jt, Alexbot, Qwfp, Bgeelhoed, Addbot, MrOllie, Ozob, Luckas~bot, Wjastle, Deepakazad, Xqbot, Bdmy, Kyng, Erik9bot, Lothar von Richthofen, Herix, Amonet, TobeBot, Trappist the monk, Jowa fan, EmausBot, User3000, AvicBot, Flatland1, ChuispastonBot, Kasirbot, Alex.j.flint, Beaumont877, MusikAnimal, Andreas27krause, Theyshallbow, Cwobeel, Dr. J. Rodal, Jochen Burghardt, NikelsenH, BeyondNormality, Parswerk, VictorM Casero, Taste wicki, Loraof and Anonymous: 61
- **Rademacher distribution** *Source:* https://en.wikipedia.org/wiki/Rademacher_distribution?oldid=677925716 *Contributors:* Michael Hardy, Tomi, Dean p foster, MisterSheik, PAR, Btyner, Entropeneur, SmackBot, Reko, Baccyak4H, R'n'B, TomyDuby, VolkovBot, DrMicro, SieBot, Qwfp, Addbot, Erik9bot, TobeBot, Duoduoduo, MidgleyC, EmausBot, ZéroBot, Koertefa, Lfj 2, QuarkyPi, ChrisGualtieri, Dexbot, BeyondNormality, Virion123, Joemore05 and Anonymous: 7
- **Binomial distribution** *Source:* https://en.wikipedia.org/wiki/Binomial_distribution?oldid=689603800 *Contributors:* AxelBoldt, Bryan Derksen, -- April, Miguel~enwiki, AdamRetchless, Youandme, Michael Hardy, David Martland, Tomi, TakuyaMurata, Ellywa, Ahoerstemeier, Charles Matthews, Timwi, Jitse Niesen, Br43402, Phr, McKay, Robbot, Sander123, Cdang, Benwing, Seglea, MichaelGensheimer, Henrygb, JB82, Robinh, Wile E. Heresiarch, Giftlite, BenFrantzDale, MSGJ, Mboverload, Knutux, LiDaobing, MarkSweep, Gauss, Atemperman, Xiao Fei, PhotoBox, Rich Farmbrough, Guanabot, Paul August, Nabla, MisterSheik, Pt, Elipongo, Musiphil, Gury, Eric Kvaalen, Atlant, Rgc-legg, PAR, Supergroupiejoy, Cburnett, Oleg Alexandrov, Postrach, Linas, Mindmatrix, LOL, Crackerbelly, Pufferfish101, Btyner, Graham87, Qwertyus, Rjwilmsi, NatusRoma, Alexb@cut-the-knot.com, Westm, New Thought, Ayla, DVdm, Volunteer Marek, YurikBot, Hede2000, Zwobot, Hirak 99, Deville, Lt-wiki-bot, Zmaboros, Ilmari Karonen, SmackBot, Blue520, Aarond10, Nbarth, Colonies Chris, Iwaterpolo, Can't sleep, clown wile eat me, Brutha~enwiki, G716, Neshatian, Rjmorris, ML5, ZantTrang, Dicklyon, Bill Malloy, Ylloh, Falk Lieder, DavidFHoughton, WeggeBot, Stebulus, Eesnyder, Janlo, Talgalili, Thijs!bot, Fisherjs, Wikid77, N5iin, Anupam, Ruber chicken, AntiVandalBot, Smachet, VectorPosse, JEH, AchatesAVC, Daytona2, MER-C, Ph.eyes, Dricherby, VoABot II, JamesBWatson, Baccyak4H, Froid, Homunq, Mmustafa~enwiki, MartinBot, R'n'B, Lucaswilkins, Mahewa, Gill110951, Coppertwig, N6ne, Spellcast, Gogobera, DrMicro, Pleasantville, Blahb31, Clay Spence, TXiKiBoT, Marvinrulesmars, Toll booth, A4bot, Steven J. Anderson, Nschuma, Stigin, Logan, WillKitch, Nguyenngaviet, Quietbritishjim, SieBot, Rlendog, BotMultichill, Gerald Tros, Garde, Hxhbot, Allmightyduck, Johnstjohn, OKBot, AlanUS, Melcombe, Redtryfan77, PsyberS, ClueBot, Rumping, Koczy, GorillaWarfare, The Thing That Should Not Be, Meisterkoch, Knassey, UKoch, Gauravm1312, Excirial, DaDexter, Watchduck, Madkaugh, Kakofonous, Qwfp, XLinkBot, Kwjbot, Efexan~enwiki, Alexius08, Tayste, MrOllie, SoSaysChappy, Tyw7, AsphyxiateDrake, Legobot, Tedtoal, Luckas~bot, Yobot, Cflm001, Qonnec, Wjastle, Moseschinyama, AnomieBOT, Erel Segal, Joule36e5, Rubinbot, MaterialsScientist, ArthurBot, Xqbot, Nasnema, J04n, Locobot, Mhadi.afraziabi, Ajs072, Citation bot 1, Intelligentsium, Wa03, Tal physdancer, Pinethicket, Stpasha, BPets, Gperjim, Mr Ape, Ian.Shannon, Tim1357, FoxBot, Sintau.tayua, TobeBot, Dinamik~bot, Bmazin, Duoduoduo, Innotata, Alzarlan16, EmausBot, Vincent Semeria, Tpudlik, Yuzisee, Welhaven, Chewings72, Sigma0

- 1, Akseli.palen, Cuttlefishy, ClueBot NG, Sealed123, Helpful Pixie Bot, Ljwsummer, BG19bot, Pallaviagarwal90, Vagobot, Mark Arsten, MC-CPO, Especially Lime, Markonius, Arr4, Dexbot, Stephan Kulla, Frosty, 069952497a, Aint one, Gmk7, Wikistiware, Hdchina2010, BeyondNormality, Sldp, Millstei, Knife-in-the-drawer, Poopedia, Will Perry, Mukhtiarhussainqazi and Anonymous: 312
- **Beta-binomial distribution** *Source:* https://en.wikipedia.org/wiki/Beta-binomial_distribution?oldid=686968336 *Contributors:* Michael Hardy, Benwing, Giftlite, Jonsafari, Rjwilmsi, Gadget850, Chris the speller, Gnp, Massbless-enwiki, Myasuda, Michael Fourman, Tomixdf, Baccyak4H, Charlesmartin14, Akiezun, Domminico, Willy.feng, Nschuma, Sheppa28, Melcombe, Thouis.r.jones, UKoch, Auntof6, Qwfp, Addbot, Luckas-bot, Yobot, Frederic Y Bois, BenzolBot, PigFlu Oink, RedBot, GoingBatty, ZéroBot, Thtanner, Jack Greenmaven, Kdisarno, Jestngrabbit, Chafe66, Sieste, Herbmuell, BeyondNormality, Bicycledreamer and Anonymous: 25
 - **Degenerate distribution** *Source:* https://en.wikipedia.org/wiki/Degenerate_distribution?oldid=664315407 *Contributors:* Bryan Derksen, Gareth Owen, XJaM, PierreAbbat, Miguel-enwiki, Michael Hardy, TakuyaMurata, Charles Matthews, PAR, Ryan Reich, Btyner, YurikBot, RussBot, Petter Strandmark, Gareth Jones, SmackBot, Radagast83, Moloch981, Baccyak4H, Quietbritishjim, Rlendog, Melcombe, Rumping, Qwfp, Addbot, MrVanBot, Legobot, Xqbot, Erik9bot, Xnn, EmausBot, Darkness Shines, IkamusumeFan, ChrisGualtieri, Bryanrutherford0, BeyondNormality, Loraof, Speinstene27 and Anonymous: 13
 - **Hypergeometric distribution** *Source:* https://en.wikipedia.org/wiki/Hypergeometric_distribution?oldid=687136612 *Contributors:* Bryan Derksen, Michael Hardy, Booyabazooka, Tomi, TakuyaMurata, Sboehringer, David Shay, Robbot, Josh Cherry, Schutz, Benwing, Giftlite, MSGJ, Zigger, Gauss, Goshng, El C, Kamrik, Maximilianh, Rgclegg, PAR, Burn, Wtmitchell, Cburnett, Linas, Mindmatrix, LOL, Pol098, Commander Keane, Btyner, Marudubshinki, RichardWeiss, YurikBot, Entropeneur, Ott2, Janto, Kingboyk, GrinBot-enwiki, Bo Jacoby, Eug, Bluebot, Nbarth, Iwaterpolo, Bilgrau, J. Finkelstein, CmdrObot, MaxEnt, Mikewax, Talgalili, Thijs!bot, PBH-enwiki, Felipehsantos, Herr blaschke, .anacondabot, Livingthingdan, Baccyak4H, David Eppstein, User A1, Antoine 245, It Is Me Here, Jia.meng, Pleasantville, Blahb31, Clay Spence, TXiKiBoT, FedeLebron, Johnlv12, Nerdmaster, Arnold90, Screech1941, Melcombe, ClueBot, Rumping, SkatingNerd, UKoch, Qwfp, Veryhuman, DavidLDill, Porejide, Alexius08, Jht4060, Addbot, DarrylNester, Wtruttschel, MrVanBot, LaaknorBot, زرشك, HerculeBot, Jack Joff, Luckas-bot, Yobot, Yvswan, AnomieBOT, Erel Segal, DirlBot, Randomactsofkindness2, Xqbot, Makeswell, ChevyC, Drcrcn, Seattle Jörg, Pröhle Tamás, Gnathan87, Duoduoduo, RjwilmsiBot, Skaphan, Peteraandrews, I9606, AManWithNoPlan, Reb42, Eidolon232, ClueBot NG, Frietjes, Gunungblau, Wbm1058, Mlxq531006, Intervallic, MC-CPO, Mtmoore321, BattyBot, Isabel duarte, Kiwi4boy, Cammy169, I3iaach, Groenger, SimonPerera, Mark viking, Equilibrium Allure, Milan Malinsky, MaEtUgR, BeyondNormality, Salubrious Toxin, Ana Caroline França and Anonymous: 115
 - **Poisson binomial distribution** *Source:* https://en.wikipedia.org/wiki/Poisson_binomial_distribution?oldid=661413838 *Contributors:* Michael Hardy, Jabowery, Rjwilmsi, Winterfors, Melcombe, Motmahp, MystBot, Addbot, Yobot, Entropeter, Rafael Calsaverini, RjwilmsiBot, ChuispastonBot, Jomtung, Yili.hong1, BeyondNormality, Monkbot, Spanachan and Anonymous: 10
 - **Fisher's noncentral hypergeometric distribution** *Source:* https://en.wikipedia.org/wiki/Fisher's_noncentral_hypergeometric_distribution?oldid=687136383 *Contributors:* Michael Hardy, RichardWeiss, SmackBot, Headbomb, SHCarter, Arnold90, UKoch, Addbot, Yobot, Citation bot, DirlBot, Nicolas Perrault III, Citation bot 1, RjwilmsiBot, CitationCleanerBot, BeyondNormality and Anonymous: 4
 - **Wallenius' noncentral hypergeometric distribution** *Source:* https://en.wikipedia.org/wiki/Wallenius'_noncentral_hypergeometric_distribution?oldid=677989858 *Contributors:* CBM, TheFearow, Magioladitis, Tomaxer, Arnold90, Melcombe, Rumping, UKoch, Yobot, Citation bot, GrouchoBot, Omnipaedista, Citation bot 1, Lucas Thoms, KLBot2, BeyondNormality, BigCrunsh-enwiki and Anonymous: 4
 - **Benford's law** *Source:* https://en.wikipedia.org/wiki/Benford's_law?oldid=690149615 *Contributors:* AxelBoldt, Calypso, Bryan Derksen, Fnielsen, ChangChienFu, Leandro, Edward, Michael Hardy, Tez, Shyamal, Tomi, Axlrosen, Ahoerstemeier, DavidWBrooks, Den fjättrade ankan-enwiki, Randywombat, Samw, Jacquerie27, Cherkash, Charles Matthews, Doradus, Taxman, BenRG, Jeffq, Gandalf61, Henrygb, KellyCoinGuy, Paul Murray, Giftlite, Wizzy, BenFrantzDale, Orangemike, MSGJ, Curps, Finn-Zoltan, Eequor, Tagishsimon, Wmahan, Mmm-enwiki, Noe, Beland, MacGyverMagic, Pmanderson, Urhixidur, Aioth, Thorwald, Ayager, Jason Carreiro, Lubaf, 4pqlinjbot, Rich Farmbrough, Smyth, Sam Derbyshire, Bender235, ESKog, Elwikipedista-enwiki, Janna Isabot, Foobaz, Jumbuck, Burn, Cburnett, Amorymeltzer, Ceyockey, Mindmatrix, Dzordzm, Waldir, Frankie1969, Reddwarf2956, RuM, Lord.lucan, Rjwilmsi, Bubba73, The wub, FlaBot, Mathbot, Harmil, Sreffler, Hillman, RussBot, Philopedia, Lsdan, Kinser, Buster79, Irishguy, Meira, Amakuha, Avraham, Eurosong, Red Jay, Carabinieri, Mrwright, Geoffrey.landis, Jack Upland, Cnglee, Groyolo, Sbyrnes321, That Guy, From That Show!, AndrewWTaylor, Das my, A13ean, SmackBot, InverseHypercube, Betacommand, Thumperward, Morte, Nbarth, Kindall, Michael.Pohoreski, Kevinpurcell, Sholom, Cybercobra, Mitar, Hgilbert, Circumspice, Derek farn, Michael Bednarek, Kompere, Vashthorvat, Hu12, DouglasCalvert, Kencf0618, IanOfNorwich, Achoo5000, Amniarix, VoxLuna, David s graff, Nunquam Dormio, Doctormatt, Reywas92, Kweeket, UncleBubba, Pcu123456789, Headbomb, Uruiamme, Escarbot, Oreo Priest, StringRay, AstroLynx, Subwaynyc, Jj137, Cyclotome, Hannes Eder, SamIAMot, JAndbot, Oxinabox, Eurobas, Mkch, PhilKnight, PChalmer, Jakob.scholbach, Baccyak4H, Paul Niquette, Johnbibby, JJ Harrison, David Eppstein, Lonewolf1313, ExoSynth, Jordi G, SteveChervitzTrutane, STEamTraen, SuneJ-enwiki, DrMicro, Pleasantville, Akwdb, PMajer, Gpeilon, Lanzkron, MBlakley, Kojones, KWRegan, Drnathanfurious, Bdb484, Cgwaldman, Sheppa28, Euryalus, Gknor, Alexxmail, BGrayson, Oxymoron83, Lightmouse, Sunrise, Sean.hoyland, Melcombe, Mr. Stradivarius, MenoBot, DavidHobby, ClueBot, Rumping, Imfranklyn, Ebster95, DragonBot, Robertharther, Sun Creator, NuclearWarfare, Qwfp, Johnuniq, Tam0031, Justin Mauger, XLinkBot, Charles Sturm, Asrghashrioadrhr, Addbot, DOI bot, Download, Protonk, LinkFA-Bot, Tassedethe, Numbo3-bot, Ehrenkater, Lightbot, Jlederluis, FarhanC99, Luckas-bot, Yobot, Nghtwlkr, Julia W, Jeremyleader, Albrodax, Azylber, AnomieBOT, Nishkarshs, Citation bot, Cactusthorn, Srich32977, CurmudgeonlyEditor, AV3000, Tarantulae, Splrzd, BYZANTIVM, Joxemai, Nameless23, Prari, Social Norm, Citation bot 1, Marcus erronius, Adlerbot, MondalorBot, Foobarnix, Trappist the monk, Gnathan87, Crtolle, Jackessler, Tbhotch, Dalba, NameIsRon, Malurth, Ash.matadeen, JaeDyWolf, K6ka, Malcolm77, U+003F, Tjifo98, Friendshao, Mathstat, Iiii I I I, Vskipper, Snotbot, Reify-tech, Bibcode Bot, Technical 13, WikiTryHardDieHard, Foxtod, TLAN38, Chafe66, Tropcho, ChrisGualtieri, Illia Connell, Dexbot, Kiwi4boy, Kkved, Lingvano, REM888, BeyondNormality, Monkbot, AmandaJohnson2014, Chuluojun, Todd Christopher Headrick, Machavanne, Petersonharryboy, Mdude2005, Meucat, Rdodds, Francesco98989899 and Anonymous: 189
 - **Beta prime distribution** *Source:* https://en.wikipedia.org/wiki/Beta_prime_distribution?oldid=672462497 *Contributors:* Michael Hardy, Tomi, Cyan, Selket, Robinh, Giftlite, Rich Farmbrough, MisterSheik, Eric Kvaalen, Bookandcoffee, Oleg Alexandrov, Btyner, Rjwilmsi, Krishnavedala, Schmock, Entropeneur, SmackBot, Maksim-e-enwiki, R'n'B, RomainThibaux, SieBot, Sheppa28, OKBot, Melcombe, Celiq, Rumping, TerryM--re, Qwfp, Mingovia, Addbot, AnomieBOT, Xqbot, RjwilmsiBot, ZéroBot, BG19bot, Purple Post-its, Herbmuell, BeyondNormality, Monkbot and Anonymous: 13

- **Birnbaum–Saunders distribution** *Source:* https://en.wikipedia.org/wiki/Birnbaum%E2%80%93Saunders_distribution?oldid=650573858 *Contributors:* Michael Hardy, Btyner, Rjwilmsi, Khazar, Magioladitis, VolkovBot, DrMicro, Melcombe, UKoch, Qwfp, Sandrobt, Addbot, Yobot, Citation bot, RjwilmsiBot, BeyondNormality, Mxalsh and Anonymous: 5
- **Chi distribution** *Source:* https://en.wikipedia.org/wiki/Chi_distribution?oldid=683998211 *Contributors:* Michael Hardy, Dbenbenn, Rich Farmbrough, MisterSheik, PAR, Btyner, Krishnavedala, SmackBot, Tom Loughheed, Aastrup, Bluebot, Nbarth, Iwaterpolo, Odedee, Harish victory, Ensign beedril, User A1, TXiKiBoT, Sheppa28, Melcombe, SchreiberBike, Qwfp, DumZiBoT, FellGleaming, Addbot, Josevellezcaldas, Meisam, Yobot, Wjastle, Erik9bot, Nicolas Perrault III, Kastchei, EmausBot, Vsbasto, ZéroBot, MelbourneStar, BeyondNormality and Anonymous: 20
- **Chi-squared distribution** *Source:* https://en.wikipedia.org/wiki/Chi-squared_distribution?oldid=690329433 *Contributors:* AxelBoldt, Bryan Derksen, The Anome, Ap, Michael Hardy, Stephen C. Carlson, Tomi, Mdebets, Ronz, Den fjättrade ankan–enwiki, Willem, Jitse Niesen, Hgambao, Fibonacci, Zero0000, AaronSw, Robbot, Sander123, Seglea, Henrygb, Robinh, Isopropyl, Weialawaga–enwiki, Giftlite, Dbenbenn, BenFrantzDale, Herbee, Sietse, MarkSweep, Gauss, Zfr, Fintor, Rich Farmbrough, Dbachmann, Paul August, Bender235, MisterSheik, O18, TheProject, NickSchweitzer, Iav, Jumbuck, B k, Kotasik, Sligocki, PAR, Cburnett, Shoefly, Oleg Alexandrov, Mindmatrix, Btyner, Rjwilmsi, Pahan–enwiki, Salix alba, FlaBot, Alvin–cs, Pstevens, Bgwhite, Philten, Roboto de Ajvol, YurikBot, Wavelength, Jtbandes, Schmock, Tony1, Zwobot, Jspacemen01–wiki, Reyk, Zvika, KnightRider–enwiki, SmackBot, Eskimbot, BiT, Afa86, Bluebot, TimBentley, Master of Puppets, Silly rabbit, Nbarth, AdamSmith, Iwaterpolo, Eliezg, Wen D House, A.R., G716, Saippukauppias, Rigadoun, Loodog, Mlgiganteus1, Qiuxing, Funnybunny, Chris53516, Tawkerbot2, Jackzhp, CBM, Rflrob, Dgw, FilipeS, Blaisorblade, Talgalili, Thijs!bot, DanSoper, Lovibond, Pabristow, MER–C, Plantsurfer, Mcorazao, J–stan, Leotolstoy, Wasell, VoABot II, Jackrystyn, User A1, TheRanger, MartinBot, STBot, Steve8675309, Neon white, Iceaturtles, It Is Me Here, TomyDuby, Mikael Häggström, Quantling, Policron, Nm420, HyDeckar, Sam Blacketer, DrMicro, LeilaniLad, Gaara144, AstroWiki, Notatoad, Johnlv12, Wesamuels, Tarkashastri, Quietbritishjim, Rlendog, Sheppa28, Phebot, Jason Goldstick, Tombomp, OKBot, Melcombe, Digisus, Volkan.cevher, Loren.wilton, Animeronin, ClueBot, Jdgilbey, MATThematical, UKoch, SamuelTheGhost, EtudiantEco, Bluemaster, Qwfp, XLinkBot, Knetlalala, MystBot, Paulginz, Fergikush, Tayste, Addbot, Fgnievinski, Fieldday–sunday, MrOllie, Download, LaaknorBot, Renatokeshet, Lightbot, Ettrig, Chaldor, Luckas–bot, Yobot, Wjastle, Johnlemartirao, AnomieBOT, Microball, MtBell, MaterialsScientist, Geek1337–enwiki, EOBarnett, DirlBot, LilHelpa, Lixiaoxu, Xqbot, Eliel Jimenez, Etoombs, Control.valve, NocturneNoir, GrouchoBot, RibotBOT, Entropeter, Shadowjams, Griffinofwales, Constructive editor, FrescoBot, Tom.Reding, Stpasha, MastiBot, Gperjim, Fergusq, Xnn, RjwilmsiBot, Kastchei, Alph Bot, Wassermann7, Markg0803, EmausBot, John of Reading, Yuzisee, Dai bach, Pet3ris, U+003F, Zephyrus Tavvier, Levdtrotsky, ChuispastonBot, Emilpohl, Brycehughes, ClueBot NG, BG19bot, Analytics447, Snouffy, Drhowey, Dlituiev, Minsbot, Dexbot, HelicopterLlama, Limit-theorem, Ameer diaa, Idoz he, Zjbranson, DonaghHorgan, Catalin.ghervase, BeyondNormality, Monkbot, Alakzi, Bderrett, Uceeylu, Michaelg2015, Zcollvee and Anonymous: 247
- **Dagum distribution** *Source:* https://en.wikipedia.org/wiki/Dagum_distribution?oldid=680740995 *Contributors:* Michael Hardy, Nabla, Rgdboer, SmackBot, KylieTastic, Sheppa28, Melcombe, Addbot, MarkAHershberger, Csigabi, Shadowjams, ZéroBot, GonzoEcon, Solomon7968 and BeyondNormality
- **Exponential distribution** *Source:* https://en.wikipedia.org/wiki/Exponential_distribution?oldid=689861550 *Contributors:* AxelBoldt, CYD, Bryan Derksen, Taw, Taral, Enchanter, Isis–enwiki, Edward, Michael Hardy, Rp, Dcljr, Tomi, Cyp, Den fjättrade ankan–enwiki, Smack, Dcoetzee, Fvw, Wilke, Robbot, Benwing, Henrygb, Decrypt3, Giftlite, Zaha, MarkSweep, Gauss, Karl–Henner, Rich Farmbrough, Vsmith, Paul August, MisterSheik, Mdf, Mwanner, Kappa, PAR, Cburnett, Jheald, Wyatts, Woohookitty, Mindmatrix, Igny, LOL, Btyner, A3r0, MekaD, Rjwilmsi, Pruneau, Mathbot, Shaile, Bgwhite, YurikBot, RobotE, Zeno of Elea, Avraham, Johndburger, Closedmouth, MStaw, Nothlit, Ilmari Karonen, Zvika, Shingkei, Zeycus, Cazort, MclD, Remohammadi, TimBentley, Iwaterpolo, Memming, Aldaron, Spartanfox86, Matroberts, CapitalR, Markjoseph125, Erzbischof, Skittleys, Talgalili, Thijs!bot, Hsne, Headbomb, JAnDbot, IanOsgood, Beaumont, .anacondabot, Coffee2theorems, A.M.R., User A1, R'n'B, Grinofadrunkwoman, Policron, Largoplazo, Jester7777, VolkovBot, Dudubur, TXiKiBoT, A4bot, Rei–bot, Z.E.R.O., Groceryheist, SieBot, Rlendog, Sheppa28, Jason Goldstick, Aiden Fisher, OKBot, Zxxterry, Water and Land, Anchor Link Bot, Melcombe, GorillaWarfare, The Thing That Should Not Be, WDavis1911, UKoch, Thegeneralguy, DragonBot, Skbkekas, SchreiberBike, Qwfp, Mejjem, MystBot, Addbot, Drevicko, MrOllie, Butchbrody, HerculeBot, Luckas–bot, Yobot, Wjastle, Kan8eDie, AnomieBOT, Ularevalo98, Sergey Suslov, ArthurBot, Bdmy, Isheden, NORbeck, Kyng, FrescoBot, BenzolBot, Oysindi, Calmer Waters, Avabait, Stpasha, Sss41, Amonet, ActivExpression, TobeBot, Duoduoduo, Kastchei, Wassermann7, Yuzisee, Yoyod, ZéroBot, Aria802, Zephyrus Tavvier, Scortchi, ClueBot NG, Asitgoes, BarrelProof, Mpaa, Helpful Pixie Bot, CD.Rutgers, Boriaj, Hyoseok, Trombonechamp, René Vápeník, ChrisGualtieri, Mogism, Burzuchius, SFK2, Linuxjava, Tertius51, Carloslizarzagac, BeyondNormality, Jodawill, Garthtarr, 16Gred, GVpep, Engheta, Melanie.zbrooks, Thefeudalspirit, Palavian, Uceeylu and Anonymous: 195
- **F-distribution** *Source:* <https://en.wikipedia.org/wiki/F-distribution?oldid=688771796> *Contributors:* Bryan Derksen, Fnielsen, Michael Hardy, Tomi, Mdebets, Cherkash, Dysprosia, Jitse Niesen, Robbot, Seglea, Henrygb, Robinh, Wile E. Heresiarch, Giftlite, MarkSweep, Oscar, O18, Arthena, Cburnett, Jheald, Btyner, RichardWeiss, Salix alba, Elmer Clark, Nehalem, Timholly, SmackBot, Unyoyega, Adouzy, Commander Keane bot, TedE, Markjoseph125, Irwangatot, Thijs!bot, PBH–enwiki, DanSoper, Zorgkang, JAnDbot, Hectorlamadrid, Albmont, Livingthingdan, Brenda Hmong, Jr, TomyDuby, Ged.R, Quietbritishjim, Sheppa28, OKBot, Melcombe, UKoch, Alexbot, Razorflame, Bluemaster, Qwfp, XLinkBot, The Squicks, Addbot, DarrylNester, Fgnievinski, MrOllie, SpBot, Jan eissfeldt, HerculeBot, Luckas–bot, Yobot, Ptbogourou, MaterialsScientist, Xqbot, GrouchoBot, JokeySmurf, Pinethicket, Tom.Reding, Gperjim, Amonet, Kastchei, EmausBot, ZéroBot, Ethaniel, Art2SpiderXL, Emilpohl, ClueBot NG, Asitgoes, HMsSolent, MusikAnimal, Califasuseso, IkamusumeFan, Kondormari, BeyondNormality, Monkbot, Loraof and Anonymous: 49
- **Fisher's z-distribution** *Source:* https://en.wikipedia.org/wiki/Fisher's_z-distribution?oldid=690265701 *Contributors:* Fnielsen, Duncharris, Eric Kvaalen, Stemonitis, Btyner, RichardWeiss, Rjwilmsi, Tevildo, Headbomb, VolkovBot, Hey jude, don't let me down, Sheppa28, Melcombe, SchreiberBike, Qwfp, Addbot, Bte99, Lightbot, WikiDreamer Bot, WikitanvirBot, ZéroBot, BeyondNormality, Monkbot and Anonymous: 5
- **Folded normal distribution** *Source:* https://en.wikipedia.org/wiki/Folded_normal_distribution?oldid=654482378 *Contributors:* Michael Hardy, Smalljim, PAR, Btyner, Rjwilmsi, Vossman, SamuelRiv, SmackBot, O keyes, Alaibot, Karho.Yau, MarshBot, Krowsky, VolkovBot, Jeff G., Rlendog, Melcombe, Worfolk, Qwfp, Addbot, DOI bot, Qorilla, Dannaf, Citation bot 1, ZéroBot, ClueBot NG, CitationCleanerBot, BeyondNormality, Upliftmofoe and Anonymous: 12

- **Fréchet distribution** *Source:* https://en.wikipedia.org/wiki/Fr%C3%A9chet_distribution?oldid=683744017 *Contributors:* Michael Hardy, Tomi, Arthena, Gene Nygaard, SmackBot, Xtaty-enwiki, Nutcracker, A. Pichler, VolkovBot, Rlendog, Sheppa28, Melcombe, Neznane, Rumping, Qwfp, Addbot, Yobot, AnomieBOT, Csigabi, Xqbot, FrescoBot, LucienBOT, Jonesey95, EmausBot, ZéroBot, JA(000)Davidson, Asitgoes, Amr.rs, Helpful Pixie Bot, BG19bot, Risk modeler, BeyondNormality and Anonymous: 11
- **Gamma distribution** *Source:* https://en.wikipedia.org/wiki/Gamma_distribution?oldid=686932783 *Contributors:* Bryan Derksen, Fnielsen, Michael Hardy, Tomi, Barak-enwiki, A5, Samsara, Phil Boswell, Robbot, Robbyjo-enwiki, Benwing, Gandalf61, Henrygb, Robinh, Dougg, Ambarish, Brenton, Wile E. Heresiarch, Giftlite, Paul Pogonysh, Darin, Fangz, MarkSweep, Gauss, MisterSheik, Bobo192, O18, Complex01, PAR, Vellea, Cburnett, LukeSurl, Linas, David Haslam, Smmurphy, Btyner, Jshadiah, Rjwilmsi, Hgkamath, Chobot, Zebediah49, DVdm, Bgwhite, Wavelength, Jlc46, Dobromila, Gaius Cornelius, Schmock, Zwobot, Entropeneur, Thomas stieltjes, Arthur Rubin, Erik144, Mebden, Bo Jacoby, Zvika, SmackBot, Adfernandes, Patrick.Wuechner, Mclid, Aastrup, Eug, MalafayaBot, Adam Clark, Colonies Chris, Arg, Iwaterpolo, Berland, Wiki me, Autopilot, Lambiam, Xtaty-enwiki, Qiuxing, Dicklyon, CapitalR, Freelancer685, Mjohnrussell, TestUser001, Shorespirit, Talgalili, Thijs!bot, Wikid77, PBH-enwiki, Lovibond, Pichote, Jirka6, Frobnitzem, Stephreg, Albmont, Baccyak4H, Lfcampos, Stevvers, User A1, Leyo, Cmghim925, Policron, STBotD, LoyalSoldier, Cerberus0, VolkovBot, DrMicro, Clay Spence, Asteadman, A4bot, Mundhenk, Thric3, Quietbritishjim, SieBot, Rlendog, Tommyjs, Donmegapoppadoc, RSchlicht, Jason Goldstick, Melcombe, JL-Bot, ClueBot, Dshutin, Alpapad, ClaudeLo, LSFenster, SamuelTheGhost, Shabbychef, True rover, Sun Creator, Frau K, Galapah, Qwfp, Bethb88, Sandro01, Gjnasaa, Abtweed98, MystBot, Paulginz, Tayste, Addbot, CanadianLinuxUser, MrOllie, Yobot, Wjastle, Nallimbot, Langmore, AnomieBOT, JonathanWilliford, Umpi77, Wiki5d, DirlBot, Bdmy, GrouchoBot, ChristopherKingChemist, Supergrane, Damiano.varagnolo, Entropeter, FrescoBot, Nicolas Perrault III, Narc813, Aple123, Tom.Reding, AmphBot, Apocalyptic, Plasticpork, Amonet, Dinamik-bot, Bobmath, Patrke, Bitlemon, Kastchei, ZéroBot, Quondum, SporkBot, Vminin, Mikhail Ryazanov, ClueBot NG, Mathstat, Xuehuit, Mpa, Qzxpqbp, WJVaughn3, Mich8611, Rockykumar1982, Nickfeng88, Josevbot, Solomon7968, Perspectiva8, Manoguru, Dlituiev, BattyBot, Illia Connell, AppliedMathematics, Wqzw.wqzw1, Evan Aad, Drewblasius, Rodionova.alenka, Tengyaow, Maththerkel, BeyondNormality, Mingyuanzhou, Berzoi075, Monkbob, Julesmath, Velvel2, Krzyswit2 and Anonymous: 236
- **Generalized gamma distribution** *Source:* https://en.wikipedia.org/wiki/Generalized_gamma_distribution?oldid=640669679 *Contributors:* Brenton, Jheald, Drbreznjev, Headbomb, Bdemeshev, Melcombe, Qwfp, Yobot, Mathstat, Helpful Pixie Bot, BeyondNormality and Anonymous: 2
- **Generalized Pareto distribution** *Source:* https://en.wikipedia.org/wiki/Generalized_Pareto_distribution?oldid=682671471 *Contributors:* Michael Hardy, Zickzack, Rlendog, Melcombe, UKoch, Leonsoftware, Yobot, AnomieBOT, Citation bot, Isheden, Mathstat, Helpful Pixie Bot, Moj-dadyr, Dexbot, Limit-theorem, Roboloni, Maththerkel, BeyondNormality and Anonymous: 9
- **Gamma/Gompertz distribution** *Source:* https://en.wikipedia.org/wiki/Gamma/Gompertz_distribution?oldid=667197172 *Contributors:* Michael Hardy, Dawnfire999, RHaworth, Rjwilmsi, Wikid77, MadmanBot, Melcombe, Dthomsen8, Yobot, FrescoBot, Jeffreyhokanson, BeyondNormality, Monkbob and Anonymous: 3
- **Gompertz distribution** *Source:* https://en.wikipedia.org/wiki/Gompertz_distribution?oldid=684371322 *Contributors:* Michael Hardy, Dawnfire999, Rjwilmsi, Kinu, Naraht, Tim bates, Wikid77, Melcombe, Excirial, Muro Bot, Addbot, Yobot, WikiDan61, Ptbogourou, AnomieBOT, Csigabi, LilHela, Trappist the monk, Svarul, BattyBot, Chancelade, CalculusOfVariations, BeyondNormality, Monkbob and Anonymous: 13
- **Half-normal distribution** *Source:* https://en.wikipedia.org/wiki/Half-normal_distribution?oldid=669792770 *Contributors:* Michael Hardy, Jérôme, PAR, Btyner, Iwaterpolo, Tevyeguy, User A1, VolkovBot, Rlendog, Melcombe, Qwfp, Addbot, Srw1138, Anders Sandberg, Wateenelende, Wikipelli, ZéroBot, Chire, Fjoelskaldr, CarlWesolowski, BeyondNormality, Juen and Anonymous: 14
- **Hotelling's T-squared distribution** *Source:* https://en.wikipedia.org/wiki/Hotelling's_T-squared_distribution?oldid=682569932 *Contributors:* Michael Hardy, Tomi, Cherkash, Robinh, Giftlite, MarkSweep, Bender235, 3mta3, Eric Kvaalen, PAR, Sean3000, Btyner, YurikBot, Adfernandes, Ck lostsword, JorisvS, Thijs!bot, Escarbot, RogierBrussee, R'n'B, Ged.R, Slysplace, Thefellswooper, Melcombe, Shabbychef, Qwfp, Addbot, Abocotok, LaaknorBot, Yobot, Xqbot, GrouchoBot, Kiefer.Wolfowitz, Tom.Reding, MehdiPedia, Amonet, Kastchei, EmausBot, ZéroBot, U3964057, BG19bot, Zzyxyzzz, Mgcampb, Attleboro, Gbstats, Wittawat, BeyondNormality, I am grungy and Anonymous: 13
- **Inverse Gaussian distribution** *Source:* https://en.wikipedia.org/wiki/Inverse_Gaussian_distribution?oldid=679292016 *Contributors:* Michael Hardy, Tomi, Giftlite, MisterSheik, Oleg Alexandrov, David Haslam, Jfr26, Btyner, Krishnavadala, SmackBot, Aastrup, Iwaterpolo, Memming, Oceanh, Rhfeng, Landrubek, Wikid77, Felipehsantos, LachlanA, Sterrys, Baccyak4H, User A1, Dima373, NickMulgan, DrMicro, Sheppa28, Deavik, Melcombe, Joellecohen, Alexbot, Qwfp, Abtweed98, Addbot, Yobot, Wjastle, AnomieBOT, KristjanJonasson, Vana Seshadri, Entropeter, FrescoBot, The real moloch57, RedBot, Dalba, Tjagger, ZéroBot, Batman50, Zfeinst, Braincricket, GKSmth, BG19bot, Manoguru, Dexbot, AppliedMathematics, Limit-theorem, Bluemix, BeyondNormality, Llbuaa and Anonymous: 35
- **Lévy distribution** *Source:* https://en.wikipedia.org/wiki/L%C3%A9vy_distribution?oldid=667008179 *Contributors:* Michael Hardy, SebastianHelm, Giftlite, Night Gyr, Delius, DBrane, Tsirel, Eric Kvaalen, Ynhockey, PAR, Gene Nygaard, Jfr26, Btyner, Mathbot, Krishnavadala, Gaius Cornelius, Dymorodrepanis-enwiki, Maechler, Digfareenough, SmackBot, Saihtam, Nbarth, Ligulembot, Caviare, Xcentaur, Gbellocchi, Kleveland, Thijs!bot, Wainson, Lovibond, GirasoleDE, Rlendog, Sheppa28, Melcombe, Badger Drink, MelonBot, Qwfp, Addbot, AndersBot, Tassedethe, 84user, Yobot, Csigabi, Ptrf, Kerack, GrouchoBot, FrescoBot, PyonDude, Kastchei, ZéroBot, WJVaughn3, Smarket-enwiki, Helpful Pixie Bot, AvocatoBot, Uniquejeff, JamieBallingall, BeyondNormality, Parpid and Anonymous: 19
- **Log-Cauchy distribution** *Source:* https://en.wikipedia.org/wiki/Log-Cauchy_distribution?oldid=674541101 *Contributors:* Michael Hardy, Corvi42, Rjwilmsi, A bit iffy, Racklever, Myasuda, Katharineamy, Rlendog, Melcombe, Qwfp, Addbot, Ben Ben, Citation bot, RjwilmsiBot, ZéroBot, Helpful Pixie Bot, George Ponderevo, GoldAccount, Lemnaminor, Stamptrader, BeyondNormality and Anonymous: 1
- **Log-Laplace distribution** *Source:* https://en.wikipedia.org/wiki/Log-Laplace_distribution?oldid=637638283 *Contributors:* Michael Hardy, Oleg Alexandrov, Ladislav Mecir, Alaibot, Johnlv12, Rlendog, Melcombe, Midx1004, Addbot, Luckas-bot, Yobot, Helpful Pixie Bot, BeyondNormality and Anonymous: 1
- **Log-logistic distribution** *Source:* https://en.wikipedia.org/wiki/Log-logistic_distribution?oldid=683180052 *Contributors:* John Vandenberg, Jérôme, Rjwilmsi, PeterSymonds, Chris the speller, Aztek41, Fetchcomms, DonAndre, VolkovBot, Rlendog, Sheppa28, Melcombe, Sevilledade, Qwfp, Addbot, GargoyleBot, Alexanderius, AndersBot, Yobot, Citation bot, FrescoBot, Citation bot 1, Tom.Reding, RjwilmsiBot, Asitgoes, Masssly, Helpful Pixie Bot, Solomon7968, Illia Connell, Mark viking, Ddev55, BeyondNormality and Anonymous: 6

- Log-normal distribution** *Source:* https://en.wikipedia.org/wiki/Log-normal_distribution?oldid=688668875 *Contributors:* AxelBoldt, Bryan Derksen, Michael Hardy, Tomi, Tkinias, Cherkash, Jitse Niesen, Fredrik, Schutz, Meduz, Wile E. Heresiarch, Wealawaga-enwiki, Giftlite, Bfinn, Paul Pogonyshv, Rubik-wuerfel, Urhixidur, Ta bu shi da yu, Ciemo, Dbachmann, Alue, ZeroOne, MisterSheik, Zachlipton, PAR, Pontus, Cburnett, Danhash, Evil Monkey, ^demon, Jeff3000, Btyner, RichardWeiss, Rjwilmsi, Raddick, MZMcBride, FlaBot, Jimt075-enwiki, Krishnavedala, Volunteer Marek, Nehalem, Bgwhite, YurikBot, Wavelength, Encyclops, Cleared as filed, Schmock, Cmglee, Lunch, SmackBot, Unyoyega, MclD, Nbarth, Iwaterpolo, Berland, Khukri, Autopilot, Ocatecir, Martinp23, Phoxhat, OsbornD, Mrdthree, A. Pichler, Floklk, DonkeyKong64, Jackzhp, Myasuda, NonDucor, Thijs!bot, PBH-enwiki, Pichote, Erxnmedia, IanOsgood, Sterrys, BenB4, Magioladitis, Albmont, Baccyak4H, User A1, Rmaus, Ricardogpn, Mange01, Thomasda, Lojiki, Mikael Häggström, VolkovBot, DrMicro, The Siktath, Philip Trueman, Leav, Edutabacman, Stigin, ColinGillespie, SieBot, Rlendog, Sheppa28, Gknor, Ciberelm, Acct4, Oxymoron83, Techman224, OKBot, Sairvinexx, Water and Land, Melcombe, Martarius, ClueBot, David.hilton.p, Philtime-enwiki, Biochem67, Umpi-enwiki, Qwfp, Humanengr, Skunkboy74, Porejide, Addbot, Fgnievinski, Till Riffert, MagnusA.Bot, Mdnahas, Wikomidia, ירן, Seriousme, RobertHannah89, Yobot, 2D, Wjastle, AnomieBOT, Erel Segal, Safdarmarwat, Frederic Y Bois, LilHelpa, Hxu, Isheden, Joxemai, Constructive editor, Rgbemy, LucienBOT, Hobsonlane, Nixphoeni, Gausselimining, Jonesey95, Stpasha, MondalorBot, IhorLviv, Trappist the monk, Occawen, Dinamik-bot, RjwilmsiBot, Llnr, Vincent Semeria, Dewritech, GoingBatty, DavidMCEddy, Rhowell77, Letsgoexploring, Donner60, Ashkax, Christian Damgaard, Nite1010, ClueBot NG, Asitgoes, Psorakis, P.O.E., Helpful Pixie Bot, Mishnadar, BG19bot, Fluctuator, Jan Spousta, Jetlee0618, Manoguru, BattyBot, Cyberbot II, Dexbot, Kiwi4boy, Mogism, Lbwhu, StriatumPDM, Mivus, Limit-theorem, Kondormari, Epicgenius, AndreaGerali, EJM86, Kuperov, HolgerBrandsmeier, Dmontier, BeyondNormality, Monkbob, Srijankedia, Ralequi, RESLND, GNAAWOWHM, Laubeg, Isambard Kingdom, Guilhermesalome and Anonymous: 201
- Lomax distribution** *Source:* https://en.wikipedia.org/wiki/Lomax_distribution?oldid=646099981 *Contributors:* Michael Hardy, Nabla, Headbomb, David Eppstein, Rlendog, Melcombe, UKoch, Addbot, AnomieBOT, Isheden, FrescoBot, Asitgoes, Mathstat, Purple Post-its, BeyondNormality and Anonymous: 4
- Geometric stable distribution** *Source:* https://en.wikipedia.org/wiki/Geometric_stable_distribution?oldid=681853267 *Contributors:* Michael Hardy, Myasuda, Vigyani, Rlendog, Melcombe, Qwfp, Addbot, Thehelpfulbot, SporkBot, Helpful Pixie Bot, Buffbills7701, BeyondNormality, Levkleb and Anonymous: 1
- Nakagami distribution** *Source:* https://en.wikipedia.org/wiki/Nakagami_distribution?oldid=617448036 *Contributors:* Michael Hardy, Selket, Giftlite, Oleg Alexandrov, Hgkamath, Vegaswikian, Krishnavedala, SmackBot, Alksb, Mlgiganteus1, Wafulz, Cydebot, Alaibot, Harish victory, Baccyak4H, VladimirSlavik, EightiesRocker, BusaJD-enwiki, Melcombe, Excesskurtosis, Qwfp, Addbot, AnomieBOT, EmausBot, ZéroBot, ClueBot NG, Golaleh, BeyondNormality, Monkbob and Anonymous: 13
- Pareto distribution** *Source:* https://en.wikipedia.org/wiki/Pareto_distribution?oldid=684130381 *Contributors:* AxelBoldt, Bryan Derksen, The Anome, Heron, Edward, Michael Hardy, Tomi, Glenn, BenKovitz, Phil Boswell, Robbot, Benwing, Henrygb, Giftlite, Paul Pogonyshv, SanderSpek-enwiki, Noe, Antandrus, Beland, MarkSweep, Chadernoon, Vivacissamamente, Fenice, MisterSheik, Nannen, O18, JavOs, Eric Kvaalen, PAR, Mindmatrix, David Haslam, Btyner, DaveApter, Rjwilmsi, Hgkamath, FlaBot, Bubbleboys, Carrionluggage, Clark Kent, Shell Kinney, Nowa, Avraham, Lendu, ChemGardener, SmackBot, Reedy, Melchoir, MclD, Jprg1966, Nbarth, GruzD, A. B., Iwaterpolo, Hve, Alexandros, Dreftymac, Joseph Solis in Australia, Cyberyder, Jive Dadson, Courcelles, Vyznev Xnebara, Teratomis, Talgalili, Headbomb, PBH-enwiki, WinBot, Widefox, Lovibond, Mack2, Olaf, Magioladitis, Am rods, Wpreston, User A1, Shomoita, Mange01, Policron, Tatrgel, Paintitblack ft, DrMicro, Tkmckenzie, Philip Trueman, Enigmaman, PhysPhD, Rlendog, OKBot, Water and Land, Melcombe, ClueBot, Rock soup, UKoch, LunaDeFerrari, Doobliebop, Qwfp, EdChem, Rror, Dthomsen8, SilvononBot, MystBot, Addbot, MrVanBot, Yobot, Ptboutgourou, Wjastle, AnomieBOT, Citation bot, Sergey Suslov, DirlBot, Obersachsebot, Xqbot, DSisyphBot, Isheden, Srich32977, Chuanren, Joxemai, Scalimani, Undsoweiter, Boxplot, Stpasha, Fentlehan, Trappist the monk, Msghani, Sander69, RjwilmsiBot, EmausBot, Fpoursafaei, Chaohuang, P3^1\$Problems, Cogiati, Ida Shaw, AMenteLibera, Financestudent, Mikhail Ryazanov, ClueBot NG, Asitgoes, Buenas días, Mathstat, Marsianus, J sandstrom, Probabilityislogic, Helpful Pixie Bot, Mgil83, Bibcode Bot, Purple Post-its, ElphiBot, AvocadoBot, Plutologist, Lindsayprior, Pankaj303er, Mojdadyr, Illia Connell, Smason79, Frosty, Jpaulson77, BeyondNormality, Monkbob, Danvildanvil, Lmpt, KasparBot and Anonymous: 121
- Pearson distribution** *Source:* https://en.wikipedia.org/wiki/Pearson_distribution?oldid=679009835 *Contributors:* Michael Hardy, Tomi, Jitse Niesen, Chuunen Baka, Benwing, Giftlite, MarkSweep, Rich Farmbrough, Bcat, Eric Kvaalen, PAR, Zzyzx11, Btyner, BD2412, Rjwilmsi, Martinpeter, Doc glasgow, Mathbot, Jengelh, SmackBot, Tevyeguy, Dicklyon, CBM, Karho.Yau, Jeff560, Robin S, Plasticup, Melcombe, Sun Creator, MelonBot, Qwfp, DumZiBoT, Addbot, DOI bot, Lightbot, Yobot, Xqbot, Citation bot 1, Kastchei, Ethaniel, Bibcode Bot, Illia Connell, Anrnusna, BeyondNormality, Monkbob and Anonymous: 23
- Phase-type distribution** *Source:* https://en.wikipedia.org/wiki/Phase-type_distribution?oldid=685873565 *Contributors:* Michael Hardy, Cameron Dewe, Bjcairms, Charles Matthews, Benwing, Giftlite, ChrisRuvalo, Pearle, Btyner, Graham87, Chenxlee, Epolk, Basten, Gareth Jones, SmackBot, Goolies flock, Abadpour, Skittleys, Bobblehead, LachlanA, Chaostik, Sgarahj2107, R'n'B, M-le-mot-dit, Slysplace, Donmegapoppadoc, Aiden Fisher, Melcombe, Mgrfan, Niceguyedc, Bender2k14, Muhandes, Aitias, GeorgeTh, Qwfp, Citation bot, Syngola, Foobarnix, Tim1357, JA(000)Davidson, Robbiemorrison, BSide, LagrangeX, Helpful Pixie Bot, BattyBot, Illia Connell, Dexbot, BeyondNormality, Huxiangking and Anonymous: 27
- Rayleigh distribution** *Source:* https://en.wikipedia.org/wiki/Rayleigh_distribution?oldid=688531864 *Contributors:* Zundark, Michael Hardy, Tomi, Timwi, Pigsonthewing, Giftlite, Paul August, El C, Kwamikagami, Kghose, O18, PAR, Rabarberski, Cburnett, Gene Nygaard, Btyner, Nanite, Mathbot, BjKa, Kri, Chobot, Krishnavedala, Nzbuu, Splash, Petter Strandmark, Crasshopper, KnightRider-enwiki, Jonathanwagner, Manfreed, MclD, Iwaterpolo, Can't sleep, clown will eat me, Memming, Ferminmx, LandruBek, Mrdthree, Dgianotti, Zylorian, Dougher, Deflective, Olaf, Fabometric, Ensign beedril, User A1, MartinBot, Steve8675309, Dbooksta, Anoko moonlight, OKBot, Melcombe, UKoch, Mel aad, Alexbot, PixelBot, Cadadril, Qwfp, Xavierstuvv, Addbot, Luckas-bot, Yobot, Wjastle, Amirobot, AnomieBOT, In digma, Xqbot, DSisyphBot, Control.valve, GrouchoBot, WaysToEscape, FrescoBot, Jc3s5h, Briardew, MastiBot, Bzzzzzzzster, Kastchei, Yuzisee, Mathstat, Larsrie, Skarmenadius, Davidagross, DudShan2, Cerabot-enwiki, Wavelet transformer, BeyondNormality, Teowey and Anonymous: 74
- Rayleigh mixture distribution** *Source:* https://en.wikipedia.org/wiki/Rayleigh_mixture_distribution?oldid=547965312 *Contributors:* Bearcat, Pigsonthewing, Dgianotti, Katharineamy, Melcombe and ArticlesForCreationBot

- **Rice distribution** *Source:* https://en.wikipedia.org/wiki/Rice_distribution?oldid=662647908 *Contributors:* Michael Hardy, Tomi, Phil Boswell, GreatWhiteNortherner, Giftlite, Jason Quinn, Rich Farmbrough, MisterSheik, O18, Mdd, PAR, Btyner, BD2412, Hgkamath, Mosama, Gadget850, Sbyrnes321, SmackBot, Iwaterpolo, Rogerbrent, Dicklyon, WikiSlasher, Sijbers, User A1, Steve8675309, Mange01, Ged.R, Dbooksta, Phe-bot, Melcombe, Mild Bill Hiccup, Dndung, UKoch, Auntof6, Gsharaf, Qwfp, Voice In The Wilderness, Addbot, Fgnievinski, Yobot, AnomieBOT, Xqbot, Control.valve, Aardliu, FrescoBot, Stpasha, Ofir michael, TobeBot, Lotje, Dinamik-bot, Kastchei, Yuzisee, RenameUser01302013, ThinkOutsideTheBox101, StanfordCommSci, Yclnjust, Thatstotaname, Mathstat, Helpful Pixie Bot, ServiceAT, Zvord, BeyondNormality and Anonymous: 40
- **Shifted Gompertz distribution** *Source:* https://en.wikipedia.org/wiki/Shifted_Gompertz_distribution?oldid=638037374 *Contributors:* Giftlite, Dawnfire999, Oleg Alexandrov, Koavf, Krishnavedala, Iwaterpolo, Danhoppe, Mack2, User A1, Josuechan, Melcombe, Qwfp, DOI bot, Yobot, Yonseca, Citation bot 1, Solarra, Helpful Pixie Bot, BeyondNormality and Anonymous: 8
- **Type-2 Gumbel distribution** *Source:* https://en.wikipedia.org/wiki/Type-2_Gumbel_distribution?oldid=585386070 *Contributors:* SimonP, Charles Matthews, Selket, MarkSweep, Eric Kvaalen, PAR, SMesser, Btyner, SmackBot, Rrburke, Derek farn, Trakesht, Faermi, Aiden Fisher, Philtime-enwiki, Qwfp, Addbot, MondalorBot, Penguinnerd121, Pgrinspan, Kondormari, BeyondNormality and Anonymous: 9
- **Weibull distribution** *Source:* https://en.wikipedia.org/wiki/Weibull_distribution?oldid=687881070 *Contributors:* AxelBoldt, Bryan Derksen, Zundark, Michael Hardy, Tomi, Cyan, Doradus, MH-enwiki, Gobeirne, Giftlite, Tom harrison, Stern-enwiki, Uppland, Bender235, MisterSheik, Kghose, O18, Smalljim, Craigyl44, PAR, Wtmitchell, Cburnett, David Haslam, Btyner, Agriculture, Rjwilmsi, YurikBot, RobotE, Corecode-enwiki, Anomalocaris, Gareth Jones, TDogg310, Avraham, Joanmg, Xareu bs, Darrel francis, Mebden, Sandeep4tech, SmackBot, WalNi, Diegotorquemada, Jason A Johnson, Iwaterpolo, Eliezg, Argyriou, Chlewbot, Dmh-enwiki, Corfuman, RekishiEJ, A. Pichler, Alfpooh-enwiki, Lachambre, Jforbett, Janlo, Bgamari, Prof. Frink, Pleitch, Felipehsantos, Nick Number, G Furtado, LachlanA, AntiVandalBot, Mack2, Gcm, Olaf, Homung, JJ Harrison, KenT, Edratzer, GuidoGer, Samikrc, Kpmiyapuram, Salih, Policron, Sam Blacketer, VolkovBot, Oznickr, TXiKiBoT, Rei-bot, Danielc192, Methou, Rlendog, Phe-bot, Dhatfield, OKBot, Water and Land, Melcombe, GioCM, Robertmbaldwin, ClueBot, Alexbot, Calimo, Qwfp, DumZiBoT, Saad31, MystBot, Addbot, LaaknorBot, Tassedethe, Yobot, Wjastle, AnomieBOT, Wiki5d, LilHelpa, Yanyanjun, Isheden, GrouchoBot, RibotBOT, FrescoBot, J6w5, Gausseliminering, Sławomir Biały, Citation bot 1, Stpasha, RedBot, Trappist the monk, Michel192cm, Ale And Quail, Stryd, RjwilmsiBot, Jowa fan, EmausBot, WikitanvirBot, Wikipelli, Slawekb, Fæ, JA(000)Davidson, Erianna, Rickysmithcmrp, Emilpohl, ClueBot NG, Asitgoes, WJVaughn3, Alexey Sanko, Helpful Pixie Bot, Epzsl2, BG19bot, Sriharid, Arr0008, SirEdvin, ReconditeRodent, Dough34, Pioneer colonel, BeyondNormality, Mr.khassi, GVpep and Anonymous: 113

3.2 Images

- **File:Ambox_important.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/b/b4/Ambox_important.svg *License:* Public domain *Contributors:* Own work, based off of Image:Ambox scales.svg *Original artist:* Dsmurat (talk · contribs)
- **File:Benford-physical.svg** *Source:* <https://upload.wikimedia.org/wikipedia/commons/8/82/Benford-physical.svg> *License:* Public domain *Contributors:* Transferred from en.wikipedia to Commons by User:Tam0031 using CommonsHelper. *Original artist:* Drnathanfurious at en.wikipedia
- **File:BenfordBroad.gif** *Source:* <https://upload.wikimedia.org/wikipedia/commons/d/de/BenfordBroad.png> *License:* Public domain *Contributors:* Own work *Original artist:* Sbyrnes321
- **File:BenfordNarrow.gif** *Source:* <https://upload.wikimedia.org/wikipedia/commons/c/c8/BenfordNarrow.gif> *License:* Public domain *Contributors:* Own work *Original artist:* Sbyrnes321
- **File:Benfords_law_illustrated_by_world's_countries_population.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/0/0b/Benfords_law_illustrated_by_world's_countries_population.png *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Jakob.scholbach
- **File:Bernoulli_distribution_chart.jpg** *Source:* https://upload.wikimedia.org/wikipedia/commons/d/d5/Bernoulli_distribution_chart.jpg *License:* CC BY-SA 4.0 *Contributors:* Own work *Original artist:* Runner1928
- **File:Beta-binomial_cdf.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/a/a4/Beta-binomial_cdf.png *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Nschuma
- **File:Beta-binomial_distribution_pmf.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/e/e1/Beta-binomial_distribution_pmf.png *License:* CC BY-SA 3.0 *Contributors:* Entirely my own work *Original artist:* Nschuma
- **File:Beta_distribution_pdf.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/9/9a/Beta_distribution_pdf.png *License:* CC-BY-SA-3.0 *Contributors:* ? *Original artist:* ?
- **File:Beta_prime_cdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/0/08/Beta_prime_cdf.svg *License:* CC0 *Contributors:* Own work *Original artist:* Krishnavedala
- **File:Beta_prime_pdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/1/18/Beta_prime_pdf.svg *License:* CC0 *Contributors:* Own work *Original artist:* Krishnavedala
- **File:Binomial_Distribution.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/b/b7/Binomial_Distribution.svg *License:* CC BY-SA 3.0 *Contributors:* Own work. Derived from File:BinDistApprox large.png by Xiao Fei, released under GFDL/CC-BY-SA-3.0. *Original artist:* cflm (talk)
- **File:Binomial_distribution_cdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/5/55/Binomial_distribution_cdf.svg *License:* Public domain *Contributors:* Own work *Original artist:* Tayste
- **File:Binomial_distribution_pmf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/7/75/Binomial_distribution_pmf.svg *License:* Public domain *Contributors:* Own work *Original artist:* Tayste

- **File:Biologist_and_statistician_Ronald_Fisher.jpg** Source: https://upload.wikimedia.org/wikipedia/commons/3/37/Biologist_and_statistician_Ronald_Fisher.jpg License: CC BY 2.0 Contributors: <https://www.flickr.com/photos/internetarchivebookimages/20150531109/> Original artist: Flickr commons
- **File:CDF-log_normal_distributions.svg** Source: https://upload.wikimedia.org/wikipedia/commons/4/42/CDF-log_normal_distributions.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:Cauchy_pdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/8/8c/Cauchy_pdf.svg License: CC BY 3.0 Contributors: Own work Original artist: Skbkekaskas
- **File:Chernoff_XS_CDF.png** Source: https://upload.wikimedia.org/wikipedia/en/b/ba/Chernoff_XS_CDF.png License: CC-BY-SA-3.0 Contributors: Own work Original artist: Willem
- **File:Chi-square_cdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/0/01/Chi-square_cdf.svg License: CC BY 3.0 Contributors: Own work Original artist: Geek3
- **File:Chi-square_distributionPDF.png** Source: https://upload.wikimedia.org/wikipedia/commons/2/21/Chi-square_distributionPDF.png License: Public domain Contributors: ? Original artist: ?
- **File:Chi-square_pdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/3/35/Chi-square_pdf.svg License: CC BY 3.0 Contributors: Own work Original artist: Geek3
- **File:Chi_distribution_CDF.svg** Source: https://upload.wikimedia.org/wikipedia/commons/1/15/Chi_distribution_CDF.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:Chi_distribution_PDF.svg** Source: https://upload.wikimedia.org/wikipedia/commons/3/35/Chi_distribution_PDF.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:Chi_on_SAS.png** Source: https://upload.wikimedia.org/wikipedia/en/9/96/Chi_on_SAS.png License: Fair use Contributors: Provided by SAS Institute Inc. Originally published on the SAS blog here. Original artist: ?
- **File:Commons-logo.svg** Source: <https://upload.wikimedia.org/wikipedia/en/4/4a/Commons-logo.svg> License: ? Contributors: ? Original artist: ?
- **File:Comparison_mean_median_mode.svg** Source: https://upload.wikimedia.org/wikipedia/commons/d/de/Comparison_mean_median_mode.svg License: CC BY-SA 3.0 Contributors: Own work Original artist: Cmglee
- **File:ComplementaryWalleniusNoncentralHypergeometric1.png** Source: <https://upload.wikimedia.org/wikipedia/en/f/fb/ComplementaryWalleniusNoncentralHypergeometric1.png> License: Cc-by-sa-3.0 Contributors: Own work Original artist: Arnold90 (talk) (Uploads)
- **File:Cumulative_distribution_function_of_Pareto_distribution.svg** Source: https://upload.wikimedia.org/wikipedia/commons/2/2a/Cumulative_distribution_function_of_Pareto_distribution.svg License: CC BY-SA 3.0 Contributors: Own work Original artist: Danvildanvil
- **File:Dagum_Distribution_(pdf).png** Source: https://upload.wikimedia.org/wikipedia/commons/a/a1/Dagum_Distribution_%28pdf%29.png License: CC BY-SA 3.0 Contributors: Own work Original artist: GonzoEcon
- **File:Degenerate.svg** Source: <https://upload.wikimedia.org/wikipedia/commons/2/2a/Degenerate.svg> License: CC BY-SA 4.0 Contributors: Own work Original artist: IkamusumeFan
- **File:Degenerate_distribution_PMF.png** Source: https://upload.wikimedia.org/wikipedia/commons/9/9e/Degenerate_distribution_PMF.png License: Public domain Contributors: ? Original artist: ?
- **File:Exponential_cdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/b/ba/Exponential_cdf.svg License: CC BY 3.0 Contributors: This graphic was created with matplotlib. Original artist: Skbkekaskas
- **File:Exponential_pdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/e/ec/Exponential_pdf.svg License: CC BY 3.0 Contributors: This graphic was created with matplotlib. Original artist: Skbkekaskas
- **File:F_dist_cdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/8/8e/F_dist_cdf.svg License: CC BY-SA 4.0 Contributors: Own work Original artist: IkamusumeFan
- **File:F_pdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/9/92/F_pdf.svg License: CC BY-SA 4.0 Contributors: Own work Original artist: IkamusumeFan
- **File:Fisher_iris_versicolor_sepalwidth.svg** Source: https://upload.wikimedia.org/wikipedia/commons/4/40/Fisher_iris_versicolor_sepalwidth.svg License: CC BY-SA 3.0 Contributors: en:Image:Fisher iris versicolor sepalwidth.png Original artist: en>User:Qwfp (original); Pbroks13 (talk) (redraw)
- **File:FishersNoncentralHypergeometric1.png** Source: <https://upload.wikimedia.org/wikipedia/en/2/22/FishersNoncentralHypergeometric1.png> License: Cc-by-sa-3.0 Contributors: ? Original artist: ?
- **File:FitExponDistr.tif** Source: <https://upload.wikimedia.org/wikipedia/commons/e/e4/FitExponDistr.tif> License: CC BY-SA 3.0 Contributors: Own work Original artist: Buenas días
- **File:FitFrechetDistr.tif** Source: <https://upload.wikimedia.org/wikipedia/commons/8/8f/FitFrechetDistr.tif> License: CC BY-SA 3.0 Contributors: Own work Original artist: Buenas días

- **File:FitLog-logisticdistr.tif** *Source:* <https://upload.wikimedia.org/wikipedia/commons/9/98/FitLog-logisticdistr.tif> *License:* Public domain *Contributors:* Own work *Original artist:* Buenas días
- **File:FitLogNormDistr.tif** *Source:* <https://upload.wikimedia.org/wikipedia/commons/f/f0/FitLogNormDistr.tif> *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Buenas días
- **File:FitParetoDistr.tif** *Source:* <https://upload.wikimedia.org/wikipedia/commons/f/f4/FitParetoDistr.tif> *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Buenas días
- **File:FitWeibullDistr.tif** *Source:* <https://upload.wikimedia.org/wikipedia/commons/2/29/FitWeibullDistr.tif> *License:* Public domain *Contributors:* Own work *Original artist:* Buenas días
- **File:Folded_normal_cdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/4/40/Folded_normal_cdf.svg *License:* Public domain *Contributors:* ? *Original artist:* ?
- **File:Folded_normal_pdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/0/02/Folded_normal_pdf.svg *License:* Public domain *Contributors:* ? *Original artist:* ?
- **File:Folder_Hexagonal_Icon.svg** *Source:* https://upload.wikimedia.org/wikipedia/en/4/48/Folder_Hexagonal_Icon.svg *License:* Cc-by-sa-3.0 *Contributors:* ? *Original artist:* ?
- **File:Frechet_cdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/d/dd/Frechet_cdf.svg *License:* GFDL *Contributors:* Self-made using python with numpy and matplotlib. *Original artist:* user:Arthena
- **File:Frechet_pdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/e/e0/Frechet_pdf.svg *License:* GFDL *Contributors:* Self-made using python with numpy and matplotlib. *Original artist:* user:Arthena
- **File:Gamma-KL-3D.png** *Source:* <https://upload.wikimedia.org/wikipedia/commons/8/8b/Gamma-KL-3D.png> *License:* CC BY-SA 3.0 *Contributors:*
- Transferred from en.wikipedia by Ronhjones *Original artist:* Mundhenk at en.wikipedia
- **File:Gamma-PDF-3D.png** *Source:* <https://upload.wikimedia.org/wikipedia/commons/b/b1/Gamma-PDF-3D.png> *License:* CC BY-SA 3.0 *Contributors:*
- Transferred from en.wikipedia by Ronhjones *Original artist:* Mundhenk at en.wikipedia
- **File:Gamma_Gompertz_cumulative_distribution_function.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/c/c7/Gamma_Gompertz_cumulative_distribution_function.png *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Dawnfire999
- **File:Gamma_Gompertz_probability_distribution.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/9/95/Gamma_Gompertz_probability_distribution.png *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Dawnfire999
- **File:Gamma_distribution_cdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/8/8d/Gamma_distribution_cdf.svg *License:* CC-BY-SA-3.0 *Contributors:*
- **Gamma_distribution_cdf.png** *Original artist:* Gamma_distribution_cdf.png: MarkSweep and Cburnett
- **File:Gamma_distribution_pdf.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/e/e6/Gamma_distribution_pdf.svg *License:* CC-BY-SA-3.0 *Contributors:*
- **Gamma_distribution_pdf.png** *Original artist:* Gamma_distribution_pdf.png: MarkSweep and Cburnett
- **File:Gompertz_cum_dist_nokey.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/2/2b/Gompertz_cum_dist_nokey.png *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Dawnfire999
- **File:Gompertz_distribution.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/5/54/Gompertz_distribution.png *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Svarul
- **File:Laplace_distribution_pdf.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/8/89/Laplace_distribution_pdf.png *License:* CC-BY-SA-3.0 *Contributors:* ? *Original artist:* ?
- **File:Levy0_LdistributionPDF.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/5/5d/Levy0_LdistributionPDF.svg *License:* CC0 *Contributors:* Own work *Original artist:* Krishnavedala
- **File:Levy0_distributionCDF.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/7/75/Levy0_distributionCDF.svg *License:* CC0 *Contributors:* Own work *Original artist:* Krishnavedala
- **File:Levy0_distributionPDF.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/6/6e/Levy0_distributionPDF.svg *License:* CC0 *Contributors:* Own work *Original artist:* Krishnavedala
- **File:LevyDistribution.png** *Source:* <https://upload.wikimedia.org/wikipedia/commons/5/58/LevyDistribution.png> *License:* Public domain *Contributors:* ? *Original artist:* ?
- **File:Logarithmic_scale.png** *Source:* https://upload.wikimedia.org/wikipedia/commons/8/8a/Logarithmic_scale.png *License:* Public domain *Contributors:* Transferred from fr.wikipedia to Commons by Esp2008 using CommonsHelper. *Original artist:* HB at French Wikipedia
- **File:Logcauchycdf.svg** *Source:* <https://upload.wikimedia.org/wikipedia/commons/3/3c/Logcauchycdf.svg> *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Qwfp
- **File:Logcauchypdf.svg** *Source:* <https://upload.wikimedia.org/wikipedia/commons/f/f3/Logcauchypdf.svg> *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* Qwfp
- **File:Loglogisticcdf.svg** *Source:* <https://upload.wikimedia.org/wikipedia/commons/7/75/Loglogisticcdf.svg> *License:* CC BY-SA 3.0 *Contributors:* Own work (Original text: self-made) *Original artist:* Qwfp (talk)

- **File:Loglogisticchaz.svg** Source: <https://upload.wikimedia.org/wikipedia/commons/1/1b/Loglogisticchaz.svg> License: CC BY-SA 3.0 Contributors: Transferred from en.wikipedia Original artist: Qwfp (talk) Original uploader was Qwfp at en.wikipedia
- **File:Loglogisticpdf.svg** Source: <https://upload.wikimedia.org/wikipedia/commons/e/e1/Loglogisticpdf.svg> License: CC BY-SA 3.0 Contributors: Transferred from en.wikipedia Original artist: Qwfp (talk) Original uploader was Qwfp at en.wikipedia
- **File:Mean_exp.svg** Source: https://upload.wikimedia.org/wikipedia/commons/d/d6/Mean_exp.svg License: CC BY-SA 3.0 Contributors: Own work Original artist: Erzbischof
- **File:Median_exp.svg** Source: https://upload.wikimedia.org/wikipedia/commons/c/cc/Median_exp.svg License: CC BY-SA 3.0 Contributors: Own work Original artist: Erzbischof
- **File:Nakagami_cdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/8/82/Nakagami_cdf.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:Nakagami_pdf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/d/d0/Nakagami_pdf.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:PDF-log_normal_distributions.svg** Source: https://upload.wikimedia.org/wikipedia/commons/a/ae/PDF-log_normal_distributions.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:PDF_Generalized_Pareto.svg** Source: https://upload.wikimedia.org/wikipedia/commons/6/6d/PDF_Generalized_Pareto.svg License: Public domain Contributors: self generated Original artist: [roboloni]
- **File:PDF_invGauss.svg** Source: https://upload.wikimedia.org/wikipedia/commons/a/a5/PDF_invGauss.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:PDF_of_Pareto_Distribution.svg** Source: https://upload.wikimedia.org/wikipedia/commons/f/f6/PDF_of_Pareto_Distribution.svg License: CC BY-SA 3.0 Contributors: Own work Original artist: Sam Mason
- **File:ParetoLorenzSVG.svg** Source: <https://upload.wikimedia.org/wikipedia/commons/b/be/ParetoLorenzSVG.svg> License: CC BY-SA 4.0 Contributors: Own work Original artist: Tkmckenzie
- **File:Pascal's_triangle_binomial_distribution.svg** Source: https://upload.wikimedia.org/wikipedia/commons/1/17/Pascal%27s_triangle%3B_binomial_distribution.svg License: Public domain Contributors: Own work Original artist: Watchduck (a.k.a. Tilman Piesk)
- **File:Pearson_system.png** Source: https://upload.wikimedia.org/wikipedia/commons/9/90/Pearson_system.png License: Public domain Contributors: ? Original artist: ?
- **File:Pearson_type_VII_distribution_PDF.png** Source: https://upload.wikimedia.org/wikipedia/commons/9/96/Pearson_type_VII_distribution_PDF.png License: Public domain Contributors: ? Original artist: ?
- **File:People_icon.svg** Source: https://upload.wikimedia.org/wikipedia/commons/3/37/People_icon.svg License: CC0 Contributors: OpenClipart Original artist: OpenClipart
- **File:Poisson_pmf.svg** Source: https://upload.wikimedia.org/wikipedia/commons/1/16/Poisson_pmf.svg License: CC BY 3.0 Contributors: Own work Original artist: Skbkakas
- **File:Portal-puzzle.svg** Source: <https://upload.wikimedia.org/wikipedia/en/f/fd/Portal-puzzle.svg> License: Public domain Contributors: ? Original artist: ?
- **File:Probability_density_function_of_Pareto_distribution.svg** Source: https://upload.wikimedia.org/wikipedia/commons/1/11/Probability_density_function_of_Pareto_distribution.svg License: CC BY-SA 3.0 Contributors: Own work Original artist: Danvildanvil
- **File:Question_book-new.svg** Source: https://upload.wikimedia.org/wikipedia/en/9/99/Question_book-new.svg License: Cc-by-sa-3.0 Contributors: Created from scratch in Adobe Illustrator. Based on Image:Question book.png created by User:Equazcion Original artist: Tkgd2007
- **File:Rayleigh_distributionCDF.svg** Source: https://upload.wikimedia.org/wikipedia/commons/a/a9/Rayleigh_distributionCDF.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:Rayleigh_distributionPDF.svg** Source: https://upload.wikimedia.org/wikipedia/commons/6/61/Rayleigh_distributionPDF.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:Rice_distribution_motivation.svg** Source: https://upload.wikimedia.org/wikipedia/commons/a/a3/Rice_distribution_motivation.svg License: CC0 Contributors: Own work Original artist: Sbyrnes321
- **File:Rice_distributiona_CDF.png** Source: https://upload.wikimedia.org/wikipedia/commons/4/41/Rice_distributiona_CDF.png License: CC-BY-SA-3.0 Contributors: ? Original artist: ?
- **File:Rice_distributiona_PDF.png** Source: https://upload.wikimedia.org/wikipedia/commons/a/a0/Rice_distributiona_PDF.png License: CC-BY-SA-3.0 Contributors: ? Original artist: ?
- **File:Rozklad_benforda.svg** Source: https://upload.wikimedia.org/wikipedia/commons/4/46/Rozklad_benforda.svg License: Public domain Contributors: Own work Original artist: Gknor
- **File:Shiftedgompertz_distribution_CDF.svg** Source: https://upload.wikimedia.org/wikipedia/commons/1/17/Shiftedgompertz_distribution_CDF.svg License: CC0 Contributors: Own work Original artist: Krishnavedala
- **File:Shiftedgompertz_distribution_PDF.svg** Source: https://upload.wikimedia.org/wikipedia/commons/0/07/Shiftedgompertz_distribution_PDF.svg License: CC0 Contributors: Own work Original artist: Krishnavedala

- **File:SkellamDistribution.png** Source: <https://upload.wikimedia.org/wikipedia/commons/b/b2/SkellamDistribution.png> License: Public domain Contributors: ? Original artist: ?
- **File:Text_document_with_red_question_mark.svg** Source: https://upload.wikimedia.org/wikipedia/commons/a/a4/Text_document_with_red_question_mark.svg License: Public domain Contributors: Created by bdesham with Inkscape; based upon Text-x-generic.svg from the Tango project. Original artist: Benjamin D. Esham (bdesham)
- **File:Tukey_anomaly_criteria_for_Exponential_PDF.png** Source: https://upload.wikimedia.org/wikipedia/commons/b/ba/Tukey_anomaly_criteria_for_Exponential_PDF.png License: CC BY-SA 3.0 Contributors: Own work Original artist: Carlos Lizarraga-Celaya
- **File:Two_red_dice_01.svg** Source: https://upload.wikimedia.org/wikipedia/commons/3/36/Two_red_dice_01.svg License: CC0 Contributors: Open Clip Art Library Original artist: Stephen Silver
- **File:Uniform_distribution.svg** Source: https://upload.wikimedia.org/wikipedia/commons/c/c2/Uniform_distribution.svg License: CC BY-SA 3.0 Contributors: Own work Original artist: Ben Moore
- **File:Wald_Distribution_matplotlib.jpg** Source: https://upload.wikimedia.org/wikipedia/commons/4/4f/Wald_Distribution_matplotlib.jpg License: CC BY-SA 3.0 Contributors: Own work Original artist: Bluemix
- **File:WalleniusNoncentralHypergeometric1.png** Source: <https://upload.wikimedia.org/wikipedia/en/f/fe/WalleniusNoncentralHypergeometric1.png> License: Cc-by-sa-3.0 Contributors: ? Original artist: ?
- **File:WalleniusNoncentralHypergeometricRecursion1.png** Source: <https://upload.wikimedia.org/wikipedia/en/5/59/WalleniusNoncentralHypergeometricRecursion1.png> License: Cc-by-sa-3.0 Contributors: ? Original artist: ?
- **File:Weibull_CDF.svg** Source: https://upload.wikimedia.org/wikipedia/commons/7/7e/Weibull_CDF.svg License: CC BY-SA 3.0 Contributors: Own work Original artist: Calimo, after Philip Leitch.
- **File:Weibull_PDF.svg** Source: https://upload.wikimedia.org/wikipedia/commons/5/58/Weibull_PDF.svg License: CC BY-SA 3.0 Contributors: Own work, after Philip Leitch. Original artist: Calimo
- **File:Wiki_letter_w_cropped.svg** Source: https://upload.wikimedia.org/wikipedia/commons/1/1c/Wiki_letter_w_cropped.svg License: CC-BY-SA-3.0 Contributors:
- Wiki_letter_w.svg Original artist: Wiki_letter_w.svg: Jarkko Piiroinen
- **File:Wikibooks-logo-en-noslogan.svg** Source: <https://upload.wikimedia.org/wikipedia/commons/d/df/Wikibooks-logo-en-noslogan.svg> License: CC BY-SA 3.0 Contributors: Own work Original artist: User:Bastique, User:Ramac et al.
- **File:Youngronaldfisher2.JPG** Source: <https://upload.wikimedia.org/wikipedia/commons/a/aa/Youngronaldfisher2.JPG> License: Public domain Contributors: <https://www.adelaide.edu.au> Original artist: Unknown

3.3 Content license

- Creative Commons Attribution-Share Alike 3.0